Faculty of Science, Technology, Engineering and Mathematics
M140 Introducing statistics

Unit 7 Factors affecting reading

Contents

Introduction

In Unit 6, we met some statistical techniques which enabled us to compare the truancy rate in large secondary schools in the East of England with the general truancy rate experienced by all secondary schools in the East of England. The method employed was to analyse sample data and use the results of the analysis to make inferences about the population from which the sample was drawn. In particular, we saw how the sign test enabled us to decide whether or not to reject, at the 5% significance level, the hypothesis that the population median takes a particular value.

Medians are just one measure of location. In this unit we return to hypotheses about location, and you will meet another hypothesis test, the 'z-test', which concerns *means*, rather than medians. For much of the unit we will be dealing with situations where we have a sample from a single population, as in Unit 6. We will then develop the ideas of hypothesis testing so as to compare *two populations* in terms of their locations. This involves setting up a hypothesis about the locations of the two populations (means, here, rather than medians) – the most common hypothesis is that the locations of the two populations are equal. A random sample of data is taken from each population, and these data are analysed to see whether or not to reject the hypothesis. Such tests are called 'two-sample tests', in contrast to 'one-sample tests' in the case of one population.

The emphasis will be on the development of statistical techniques, and, as in Unit 6, we shall explore many of the ideas in the context of a question taken from the general area of education. This time we shall be looking at the achievement of 7- and 8-year-old children in reading:

What factors affect a child's reading ability?

Section 1 starts with a brief discussion of this question. We shall then look at an available source of data and identify what aspects of the general question we can consider.

The next step will be to define specific questions of interest and use them to set up appropriate hypotheses. We will then begin to develop an appropriate sample statistic – a test statistic – with which to perform our hypothesis tests, by revisiting the idea of sampling distributions in Section 2. This notion will lead us to consider a particular distribution known as the 'normal distribution'. In Section 3, we look closely at this distribution, which is of great importance in statistics.

You were introduced to sampling distributions in Unit 4; they were used somewhat implicitly in Unit 6.

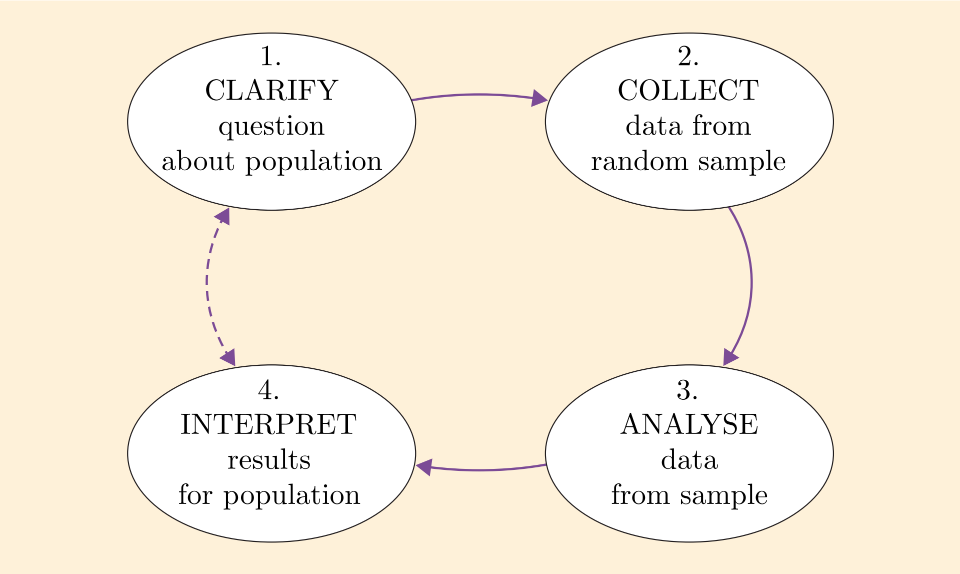
In Section 4, we go on to consider how the normal distribution helps us to define a usable test statistic, along with its sampling distribution. Section 5 is concerned with the application of the resulting z-test to the analysis of a sample of data from one population. Section 6 extends these ideas to investigate the difference between the means of two populations. One important aspect of these z-tests is that they are suitable only for dealing with (quite) large samples of data.

In Section 7, you will use Minitab to perform z-tests and learn to interpret the resulting p-values. Section 8 draws some conclusions about the educational question raised in the first section, and makes some general points about z-tests.

Section 7 directs you to the Computer Book. You are also guided to the Computer Book at the end of Subsections 3.1 and 3.3.

1 Clarifying the question

In Unit 6, the modified modelling diagram was introduced.



The modified modelling diagram (Figure 2 from Unit 6)

In this section, you are going to consider the first two stages of the modified modelling process: *clarify question* and *collect data*.

1.1 The question to be clarified

The question *What factors affect a child's reading ability?* is rather too general for us to attempt to answer it straight away. We need to make it more explicit. The first step is to understand what is meant by *reading ability*.

Activity 1 Measuring reading ability

How would *you* measure reading ability? Write down two or three measures of reading ability of 7- and 8-year-old children that you might use.

There are various different reading tests available to teachers, and they normally combine several of the measures mentioned in the solution to Activity 1. We shall be using data that have already been collected for us, so the measures used have already been defined.

A reading class

The next step is to consider the factors that might affect a child's reading ability.

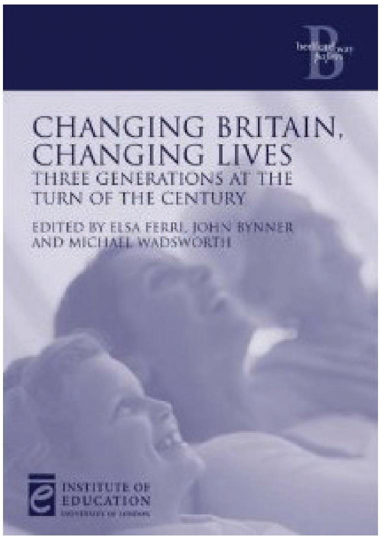
Activity 2 Factors affecting a child's reading ability

Write down some factors that you think might affect a child's ability to read.

The data that we shall use to explore this area will not allow exploration of all these factors. Therefore the data have to be examined before a decision can be made as to which factors can be explored and what questions can be asked.

1.2 The data to be used

We shall be looking at the population of British children aged 7 and 8. The sample we shall use consists of 7- and 8-year-old children of a certain group of parents defined as follows. At least one of the child's parents – a 'cohort member' – was born in a particular week in April 1970, resides in Great Britain, and has been part of a long-term study known as the British Cohort Study (BCS). There were more than 17 000 such people.



A book based on the results of the BCS and similar studies

The BCS had its origins in what was called the British Births Survey, which was originally designed to examine the social and biological characteristics of the cohort member's mother. That study looked at neonatal morbidity, and its results were compared with those of a similar, earlier study, the National Child Development Study, carried out in 1958. 'Neonatal morbidity' refers to disease of the child (the cohort member) in its first month of life.

Since 1970, the aims of the BCS have broadened considerably. There have been eight follow-up surveys, or 'sweeps', carried out in 1975, 1980, 1986, 1996, 1999–2000, 2004–2005, 2008–2009 and 2012. The follow-up surveys attempted to trace the original sample and, in the case of the first two follow-ups, to include immigrants born in the same week as the original sample. Each follow-up survey looked at different areas of the original group's development into adulthood. The one which included reading skills of the cohort member's children, and from which the data we shall be using have been taken, was the one carried out by the Centre for Longitudinal Studies at the Institute of Education, University of London, in 2004–2005. At that time, the age of the people under study was 34 years.

We are therefore going to concentrate on data relating to the reading ability of children who were 7 and/or 8 years old in 2004–2005 and whose parents were part of the BCS. The 2004–2005 sweep of the BCS provided data on 745 children aged 7 or 8 in total. Of these, only 679 were tested for their reading ability. It is this sample of 679 that we shall concentrate on in this unit. As we progress through the analysis, you will find that we shall be using sample sizes smaller than this, because not all the additional information needed was provided in the answers to the questionnaire. However, the sample sizes concerned will remain pretty large.

BAS II was, in fact, updated to BAS 3 in 2011.

Activity 3 Is it a random sample?

Write down some reasons why this sample of children can or cannot be considered a random sample of the population of 7- and 8-year-old children in Great Britain in 2004–2005.

We shall return to the issue of randomness of the sample in Section 8, but for most of the unit, despite our doubts, we shall assume that it *is* acceptable to treat the sample as if it were a random sample.

We next need to consider what data we have available. Table 1 shows data on the first few children in the sample. In the first column is the child's reading ability as scored using a standard reading test called the BAS II Word Reading Ability Score, where BAS stands for 'British Ability Scales'. This value will be referred to simply as the child's 'reading score' in this unit. The second column gives the child's age in months.

The remaining columns of Table 1 are in coded form; that is, they use simple numerically coded values to represent attributes of the child in place of more

complicated wordings, ranges of numbers or exact numerical values. For example, the third column shows the gender of the child, coded as 1 for a boy, 2 for a girl. The fourth column again relates to age; this time whether the child is aged 7 or 8 is recorded. The fifth column, headed 'Parental education', actually shows whether the cohort member's partner/spouse finished full-time education by the age of 16 or at some age over 16; it is used here as a measure of the level of education of the child's parents. The sixth column shows the occupation of the child's father. The codes for the values in columns three to six are given beneath the main body of the table.

Table 1 Part of the dataset on reading from BCS 2004–2005

Reading score	Age (months)	Gender	Coded age	Parental education	Father's occupation
106	91	1	1	1	–
123	95	2	1	1	1
123	86	2	1	–	1
110	92	1	1	1	1
92	90	2	1	2	–
129	93	2	1	1	–
118	97	1	2	–	2
115	107	2	2	1	2
117	93	2	1	2	–
134	89	1	1	1	1
25	85	1	1	–	2
110	93	2	1	1	1
172	94	1	1	1	1
138	90	2	1	–	2
56	105	1	2	–	1
136	100	2	2	1	1
115	90	2	1	1	–
160	94	1	1	1	1
⋮	⋮	⋮	⋮	⋮	⋮

'Gender', 1: boy; 2: girl. 'Coded age', 1: 7 years old; 2: 8 years old. 'Parental education', 1: finished aged 16 or less; 2: finished aged over 16. 'Father's occupation', 1: managerial, technical, professional and skilled non-manual occupations; 2: skilled manual, partly skilled and unskilled occupations.

You will notice that in some cases information is missing in the sample data. This is to be expected, because some people either cannot or do not wish to answer specific questions in the questionnaire. The missing data will just be ignored for now, but we will return to a brief consideration of its possible effects in Section 8.

A number of factors may have an effect on a child's reading ability. With our choice of data, the factors we can consider are child's age, child's gender, parental education and father's occupation.

Consider 'child's gender' first. What precise question can we ask? It must, as usual, be about the appropriate population (that of British children aged 7–8 in 2004–2005) and not merely the sample. We might ask, *Within this population, do boys and girls differ in their reading ability?* But we should be more precise. As in Unit 6, we shall be looking for a difference in *location*, in this case between reading scores of boys and girls.

In the next subsection we shall be more precise about the particular measure of location to use, but for now a reasonably precise question is:

> *For British children aged 7–8 in 2004–2005, did boys' and girls' ability scores differ in location?*

Similar questions can be asked about the other factors.

> *For British children aged 7–8 in 2004–2005, did reading scores differ in location according to their father's occupation?*

The questions can also be made more focused. For instance, consider the first question again. Perhaps there is a difference between boys and girls aged 7, but no such difference for 8-year-olds. Because of possibilities like this, it may be more appropriate to consider the two age groups separately, asking

> *For British children aged 7–8 in 2004–2005, did boys' and girls' ability scores differ in location?*

as well as

> *For British children aged 7–8 in 2004–2005, did reading scores differ in location according to the level of parental education?*

> *For British children aged 7–8 in 2004–2005, did reading scores differ in location according to their father's occupation?*

The questions on education and occupation could also be split up according to age, in a similar way.

1.3 Setting up the hypotheses

We shall try to answer most of these very specific questions by means of hypothesis tests. Let us then remind ourselves what is involved by referring back to the procedure for the sign test discussed in Unit 6, but setting it up in a more formal way.

We began by making a statement about the population of interest that we wished to test. In particular, this was the hypothesis that the population median was equal to a specified value, M. The hypothesis that the population median is equal to M is known as the **null hypothesis**. This hypothesis is usually denoted by the symbol H_0. Thus the null hypothesis in the case of the sign test can be stated precisely in the form

H_0: Population median $= M$.

We then looked at the data to see if there was any evidence that the population median did not, in fact, equal M. If there is evidence against the truth of the null hypothesis, H_0, then we *reject* this hypothesis and we conclude that there is evidence that the population median is not equal to M. That the population median is not equal to M is called the **alternative hypothesis**. An alternative hypothesis is usually denoted by the symbol H_1. Thus, if we reject the null hypothesis

H_0: Population median $= M$,

'$\neq$' is the symbol for 'is not equal to'.

then we are left with the alternative hypothesis

H_1: Population median $\neq M$,

and we say we are rejecting the null hypothesis *in favour of* the alternative hypothesis.

In Unit 6, a trial in a law court was used as an analogy to hypothesis testing. In that context, the null hypothesis is that 'the defendant is not guilty', while the alternative hypothesis is that 'the defendant is guilty'. If the evidence against the

null hypothesis is sufficiently great, then the jury should reject that hypothesis in favour of the alternative hypothesis, and conclude that the defendant is guilty.

Returning to the questions concerning children's reading ability, the first step therefore is to set up the appropriate null and alternative hypotheses. As you might expect, these correspond to whether or not there is a difference in location in reading ability between two groups of children. We shall start with one of the questions on gender.

> *For British children aged 7 in 2004–2005, did boys' and girls' reading scores differ in location?*

However, before defining the hypotheses, it is worth thinking again about the data. We actually have data on reading scores and gender for 396 children who are aged 7 (of these, 206 are boys and 190 are girls). Printing all 396 scores here would clearly be cumbersome and waste space. We can summarise the data as shown in Table 2.

Table 2 Summary statistics for data on reading scores of 7-year-old children

	Sample size	Sample mean	Sample standard deviation
Boys	206	109.31	27.671
Girls	190	113.42	25.464

(This data is copyright and owned by the Economic and Social Data Service.)

You may well be wondering why the summary measures are the mean, $\overline{x}$, and standard deviation, s, and not some other measures of location and spread, such as the median, M, and interquartile range, IQR. A minor reason is that the mean and standard deviation are commonly used in practice, so more people are familiar with them than with other measures. The main reason, though, is that $\overline{x}$ and s can be used to construct a reasonably simple test, in a way that M and IQR cannot.

Activity 4 Calculating a mean and standard deviation

Because it is some time now since you worked with the sample mean ($\overline{x}$) and the sample standard deviation (s), here is a reminder of how to calculate these summary measures:

The calculation of the mean was discussed in Subsection 1.3 of Unit 2 and the calculation of the standard deviation was discussed in Subsection 3.1 of Unit 3.

$$\overline{x} = \frac{\sum x}{n},$$

where $\sum x$ is the sum of all the sample values and n is the sample size; and

$$s = \sqrt{\text{variance}},$$

where the variance is $\dfrac{\sum(x-\overline{x})^2}{n-1}$.

(a) Data on the first eighteen 7- and 8-year-olds taken from the BCS 2004–2005 results were given in Table 1 (in Subsection 1.2). Extract from that table the values of the reading scores for all the 7-year-old boys. What is the value of n for this small sample?

(b) Calculate $\overline{x}$ and s for the reading scores for 7-year-old boys that you extracted in part (a).

Having paused briefly to examine the sample data, we now move on. We still need to state the null and alternative hypotheses associated with the question

For British children aged 7 in 2004–2005, did boys' and girls' reading scores differ in location?

in their precise forms. The null hypothesis will be

H_0: For British children aged 7 in 2004–2005, the mean reading score for girls was equal to the mean reading score for boys.

As you will have noticed, H_0 is phrased in terms of the population means and not, for example, the population medians. The alternative hypothesis is naturally taken to be

H_1: For British children aged 7 in 2004–2005, the mean reading score for girls was not equal to the mean reading score for boys.

The null and alternative hypotheses for other questions listed at the end of the previous subsection are similar.

The British Ability Scales reading score system gives overall mean test scores for different age groups in Great Britain. These overall mean test scores are given for quite finely defined age groups, from which the authors of this unit have come up with the following means for 7- and 8-year-olds: the population mean for 7-year-old children is 96, and for 8-year-olds it is 116. (Actually, these means come from very large samples of children and not the whole population, but in practice we can treat them as population means.) So a further appropriate question to ask about the data on reading scores for 7-year-old children, for example, is whether they are consistent with a population mean of 96. In other words, we could test the following hypotheses:

H_0: For British children aged 7 in 2004–2005, the mean reading score was equal to 96

H_1: For British children aged 7 in 2004–2005, the mean reading score was not equal to 96.

In testing hypotheses about population medians in Unit 6, the next step was to define a quantity that we could calculate from the data that would help us to evaluate the truth or otherwise of the null hypothesis. In the law-trial analogy, this is the evidence. In the sign test, this quantity was the smaller of the numbers of [$+$]s and [$-$]s that the sample contains. (See Section 4 of Unit 6.) In general, in hypothesis testing, this quantity is called the **test statistic**. So now we need to find suitable test statistics to assess the hypotheses about children's reading abilities. Since these hypotheses are about population means or differences between population means, the obvious test statistics would involve sample means or the differences between sample means. But, as with the sign test in Unit 6, the awkward part involves finding what is called the sampling distribution of the test statistic; so in the next section we look again at sampling distributions.

Exercises on Section 1

Exercise 1 Mean and standard deviation for 8-year-olds

(a) Extract from Table 1 the values of the reading scores for all the 8-year-old children in the table. What is the value of n for this small sample?

(b) Calculate $\overline{x}$ and s for the reading scores for 8-year-old children that you obtained in part ((a)).

Exercise 2 Parental education and occupation

(a) Extract the values of the reading scores for all the children in Table 1 whose parent's age on finishing full-time education was 16 or less and whose

father's occupation is managerial, technical, professional or skilled non-manual. What is the value of n for this small sample?

(b) Calculate $\overline{x}$ and s for the reading scores for the sample of children that you obtained in part ((a)).

Exercise 3 Null and alternative hypotheses?

Suggest null and alternative hypotheses for comparing the reading abilities of the 8-year-old children according to their gender.

2 Sampling distributions revisited

In Subsection 3.3 of Unit 4, you saw what is meant by the sampling distribution of the median of a sample, and what happened to such sampling distributions as the sample size increased. We now review these ideas, but rather than just repeating exactly what was done before, we look at the **sampling distribution of the mean** as opposed to that of the median.

As in Unit 4, in order to look at these sampling distributions precisely, we really need to know all the relevant information about the whole population. Nobody has information about the reading ability of all 7- and 8-year-old children in Great Britain, so we cannot work with data exactly like those from the BCS. Instead let us look at a population where we *do* have data on everyone, and investigate sampling distributions using that. The population is that of all students taking the examination for the Open University module *Exploring mathematics* (MS221) in a particular presentation. There were 1234 students in the presentation chosen, and their marks in the examination are displayed in Figure 1. This plot is very like a histogram with lines instead of bars. The numbers of students achieving each mark from 0 to 100 are given by the heights of the lines drawn at each mark. These heights are the same as the areas of the bars that would have been used on the histogram. But, in addition, the top ends of the lines have been joined together.

Histograms were introduced in Subsection 1.5 of the Computer Book.

This representation gives a good picture of the shape of the *population distribution* of examination marks of students on MS221 in one presentation.

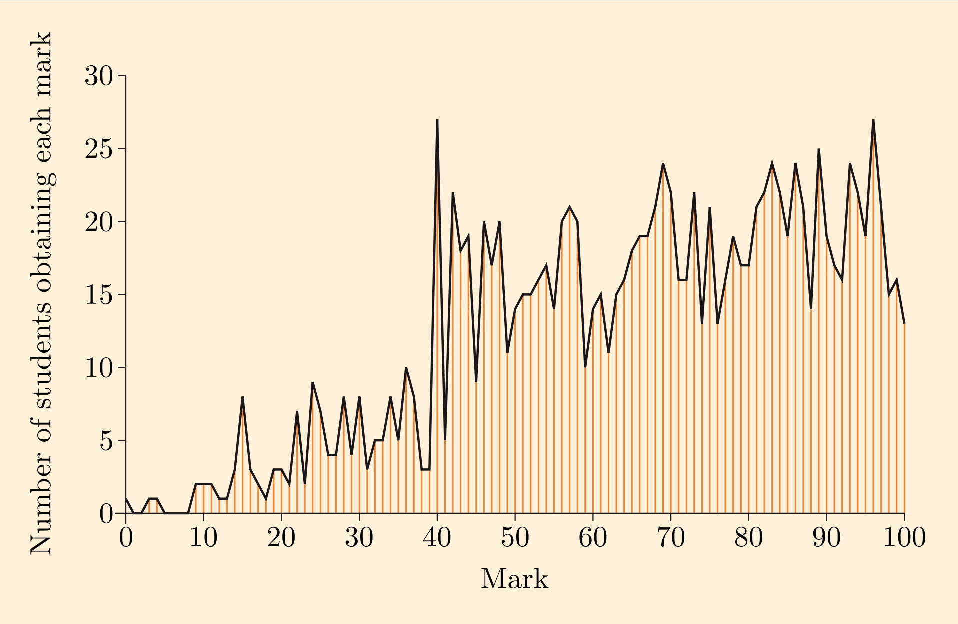


Figure 1 Numbers of students obtaining each examination mark in MS221

Now, there is a modification that we need to make. What will be important later are the *proportions* of students in the population gaining each mark. Thus instead of using the vertical axis to measure the actual number of students who obtained each mark in the examination, we want the population distribution to be described in terms of the proportion of students in the population who obtained each mark. We can do this simply by dividing each of the actual numbers represented in Figure 1 by the total number of students in the population (1234). Hence,

$$1 \text{ becomes } \frac{1}{1234} \simeq 0.0008,$$
$$2 \text{ becomes } \frac{2}{1234} \simeq 0.0016,$$
$$3 \text{ becomes } \frac{3}{1234} \simeq 0.0024,$$

and so on.

Activity 5 From number to proportion

The actual number of students scoring 75 marks in Figure 1 is 21. What proportion of students on MS221 in the presentation in question achieved 75 marks?

The result of changing from actual numbers to proportions is shown in Figure 2. Notice that Figure 2 looks just the same as Figure 1; only the scale on the vertical axis has changed.

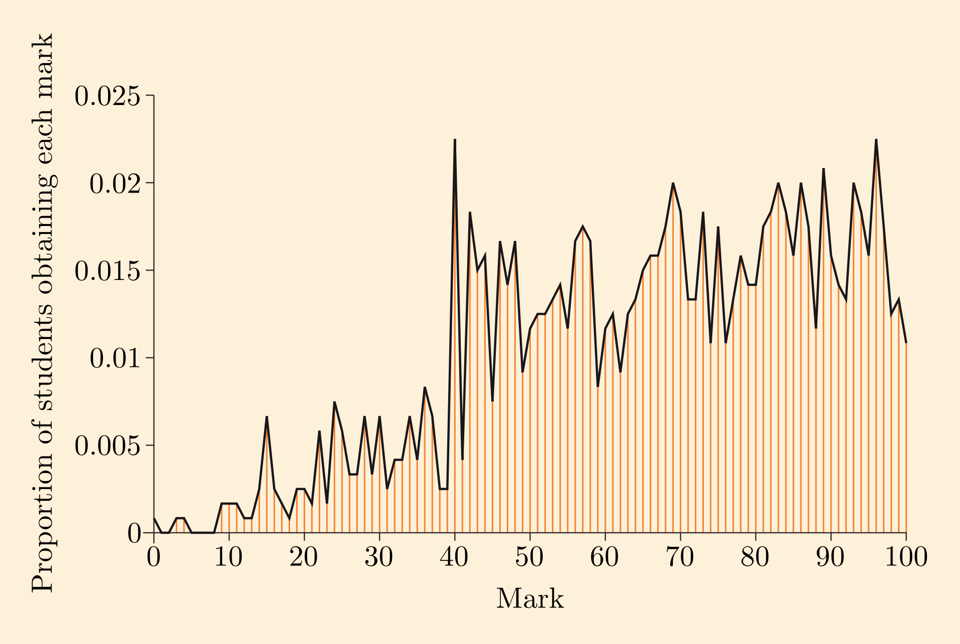


Figure 2 Proportions of students obtaining each examination mark in MS221

However, it is not the characteristics of the population distribution of exam marks, above, that we are interested in as such. Our focus is going to be on the sampling distributions of *means of random samples of exam marks taken from this population*. This is because we will be interested in testing hypotheses about

the mean examination mark, such as

H_0: For students on MS221, the mean examination score is equal to 65

H_1: For students on MS221, the mean examination score is not equal to 65

or (using data from other years)

H_0: For students on MS221, the mean examination score for the current presentation is equal to the mean examination score for the previous presentation

H_1: For students on MS221, the mean examination score for the current presentation is not equal to the mean examination score score for the previous presentation.

Now we begin our investigation of the sampling distribution of the mean. Consider first all possible random samples of size 2 that we might select from the population data of 1234 examination marks. There is a great number of possibilities (760 761, to be precise!), and we cannot concisely picture *all* the sample values in every one of these possible samples. However, as in Unit 4, we *can* summarise each sample using a *summary measure*, and then picture these in the form of the sampling distribution of that summary measure. This time, as suggested above, we use the *sample mean* as our summary measure.

Activity 6 Sample means of samples of size 2

(a) Find the sample means of each of the following samples of size 2:

(i) 15, 35 (ii) 65, 77 (iii) 65, 52 (iv) 37, 80.

(b) The exam marks in the population, and hence in any sample, are all integers (whole numbers). Are the sample means of samples of size 2 necessarily integers? If not, what other kinds of value can these sample means take?

In this way it would be possible to calculate the sample mean for every one of the 760 761 possible samples of size 2. Different samples can give the same sample mean, as (iii) and (iv) in part (a) of Activity 6 illustrate. The sampling distribution records the *proportions* of all these samples with each value of the sample mean. A picture of this is shown in Figure 3. This represents the *sampling distribution of the mean for samples of size 2* from the population of exam marks. Here all the possible values of the sample mean $\overline{x}$ are indicated on the horizontal axis, and the vertical lines represent the heights of the bars that would be used for a histogram of the proportion of samples (out of 760 761 possibilities) which have each of these values as the sample mean. Notice that there are many more lines in this diagram than there are in Figure 2. That is because this sample mean can take about twice as many values, integers and half-integers, as you saw in Activity 6, so the histogram can have twice as many bars. Joining the tops of the lines again provides us with a good picture of the *shape* of the distribution.

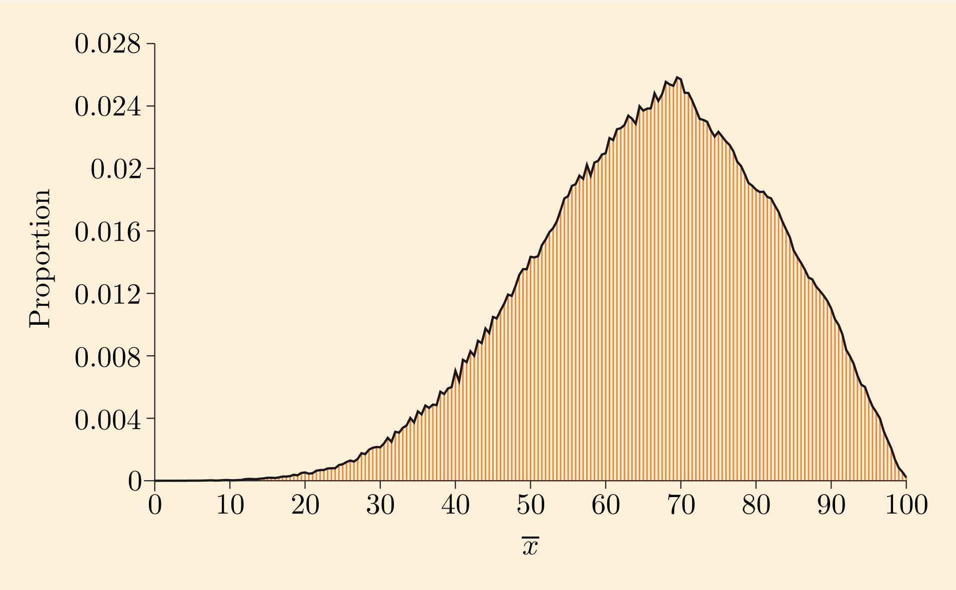


Figure 3 Sampling distribution of the mean for samples of size 2 from the population of MS221 exam marks

Activity 7 Distribution of sample means of size 2

What are the main features of the distribution of sample means of size 2 shown in Figure 3?

Let us now find out, as in Unit 4, what happens to the sampling distribution as the sample size increases. Let's first look at the sampling distribution of the mean for samples of size 3.

Activity 8 Sample means of samples of size 3

Find the sample mean of each of the following samples of size 3:

(a) 10, 20, 45 (b) 82, 24, 33 (c) 52, 61, 73 (d) 78, 64, 46.

Activity 8 indicates that there are even more possible values of the sample mean for samples of size 3 than there are for samples of size 2. This means that the vertical lines in the sampling distribution will be even closer together. For this reason, we stop plotting the lines and just concentrate on the *shape* of the distribution as indicated by the *tops* of the lines; we obtain the picture of the sampling distribution shown in Figure 4. In fact, the 'joining' line shown in Figure 4 is made up of lots of very short lines, each one joining two adjacent vertical lines.

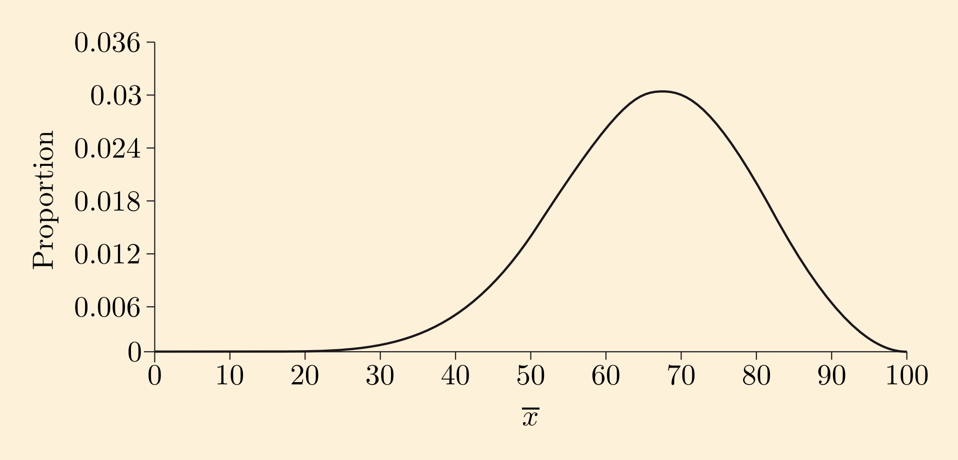


Figure 4 Sampling distribution of the mean for samples of size 3 from the population of MS221 exam marks

Activity 9 Distributions of sample means of sizes 2 and 3

How does the distribution of sample means of size 3 shown in Figure 4 compare with the distribution of sample means of size 2 shown in Figure 3?

Activity 10 Distributions of sample means of sizes 3 and 5

Figure 5 shows the distribution of sample means of size 5 from the population of MS221 examination marks. How does the distribution shown in Figure 5 compare with the distribution of sample means of size 3 shown in Figure 4?

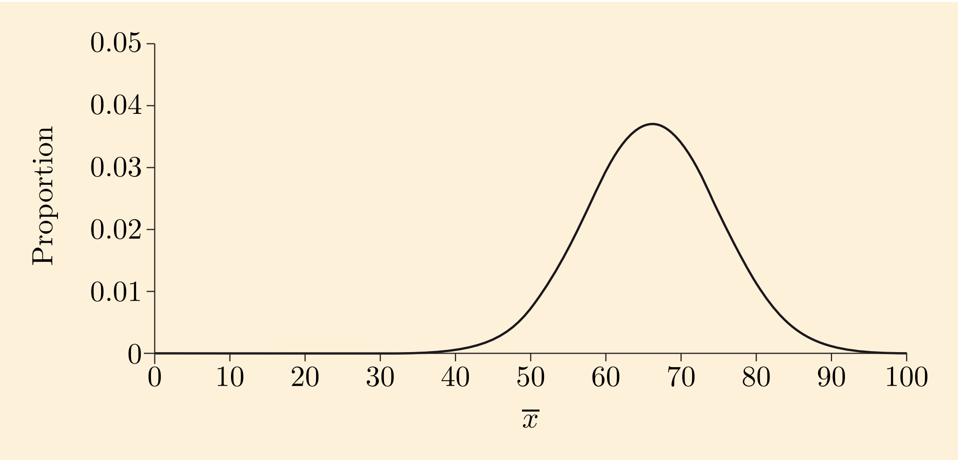


Figure 5 Sampling distribution of the mean for samples of size 5

Activity 11 Distributions of sample means of larger sample sizes

Figure 6 contains pictures of the sampling distributions of the mean for larger sample sizes. Notice that we have not indicated the scale on the vertical axes in Figure 6, but it is the same in each case. Describe the changes in shape of these distributions, as the sample size n increases.

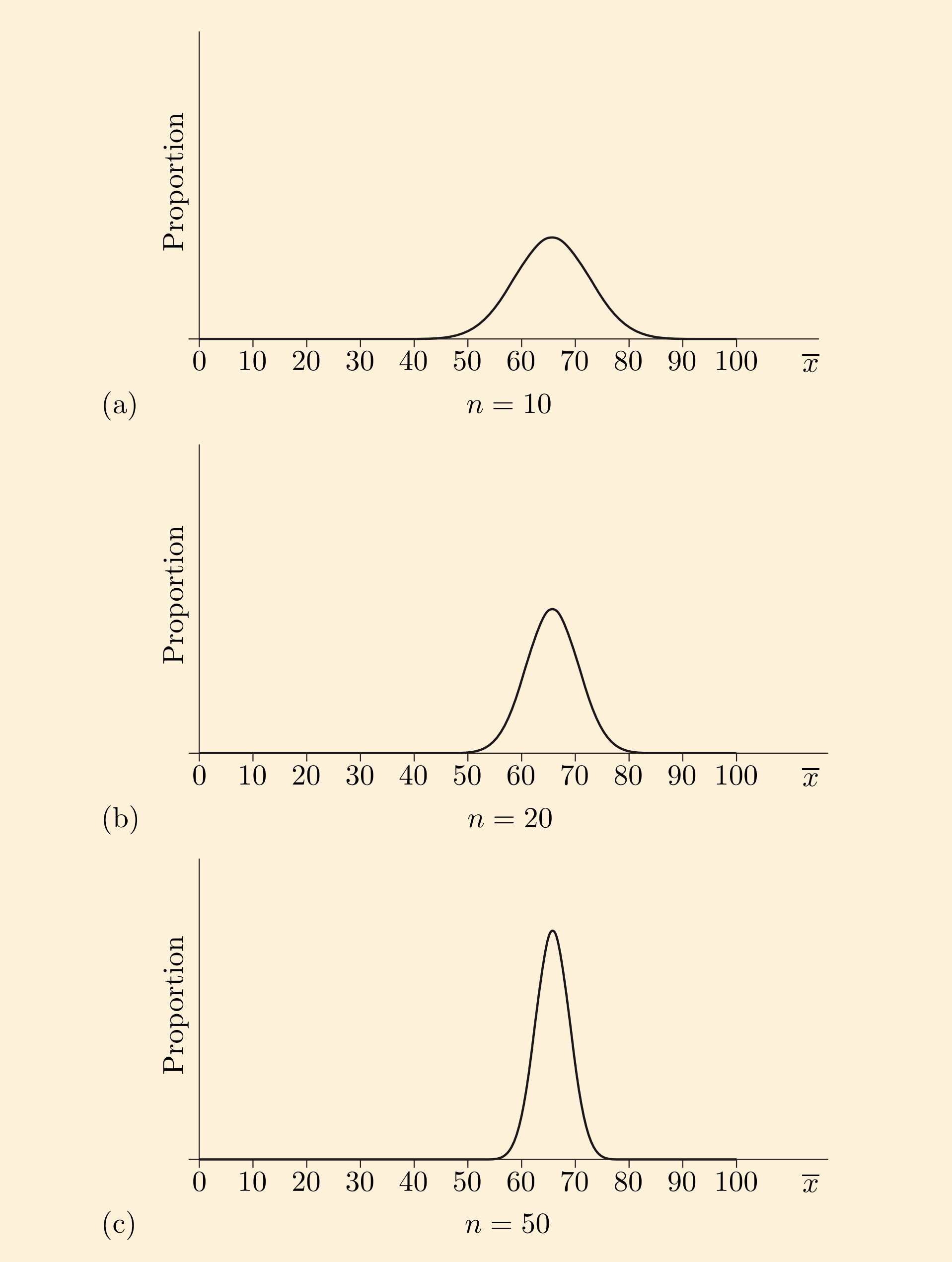


Figure 6 Sampling distributions of the mean for samples of size n

The common shape of the distributions in Figures 4, 5 and 6 is sometimes called a 'bell shape'. You can use the following picture of Big Ben to decide whether or not you agree that these distributions are bell-shaped!

Big Ben: bell-shaped?

Now the interesting thing about the sampling distribution of the mean is that it will nearly always be approximately bell-shaped (looking something like the above figures), *no matter what population distribution is taken as the starting point.* (The sampling distributions of some other quantities, such as the sample median, show similar features.)

Example 1 Sampling distributions of means based on earnings data

Figure 7 provides a rough picture of the population distribution of earnings of all full-time employees in the UK in 2011.

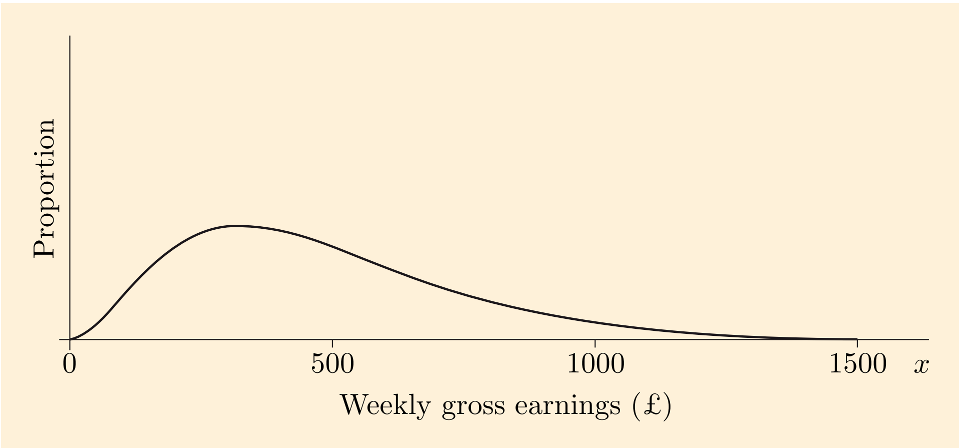


Figure 7 Population distribution of earnings of full-time employees

This population distribution is very smooth. The smoothness results from the fact that the population is extremely large and there are so many possible earnings that we can record. This means that the vertical lines representing the various adjacent proportions would be so close together that we could not distinguish between them and so, effectively, the line joining the tops is a smooth curve. The distribution is, however, clearly right-skew since it has a long tail to the right. This reflects the fact that while most employees earn a moderate to 'medium' wage, some employees earn considerably more, and a few earn very considerably more again.

Figure 8 contains pictures of the sampling distributions of the *mean* for samples of various sizes from this population distribution.

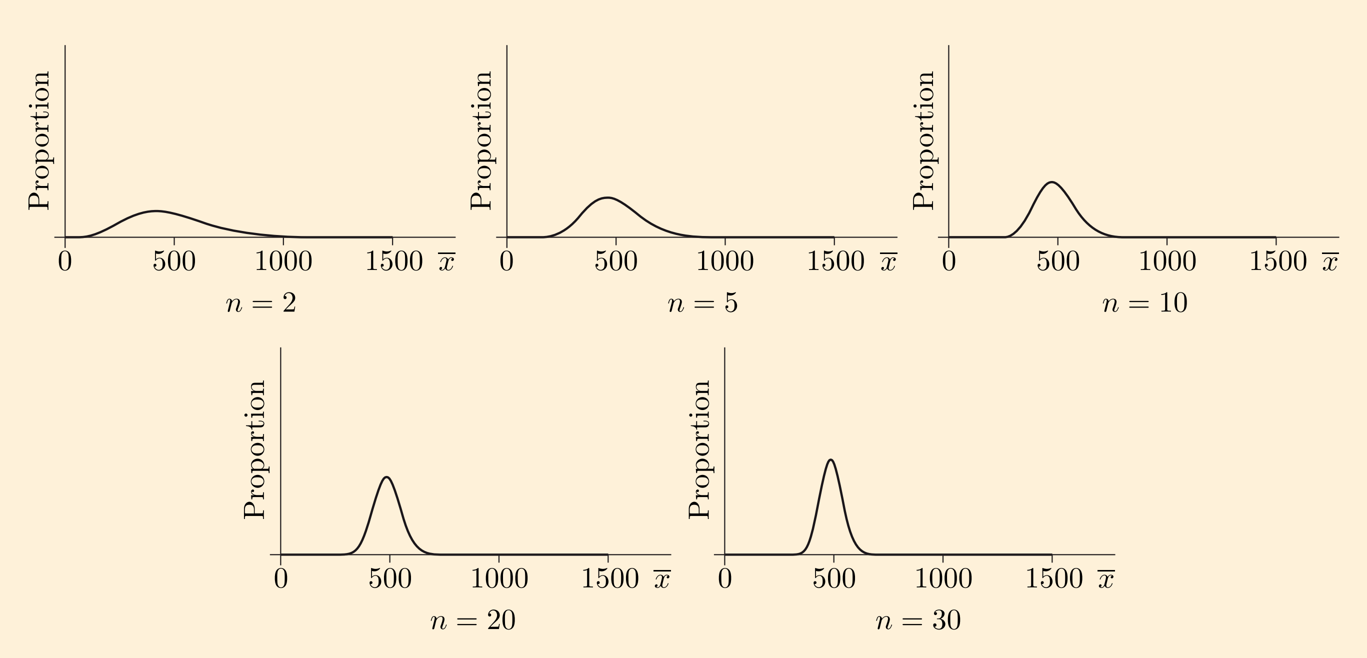


Figure 8 Sampling distributions of means based on earnings data

Activity 12 Distributions of sample means of earnings data

Describe the main changes in shape of the sampling distributions in Figure 8, as the sample size n increases.

So, again, we see from Example 1 and Activity 12 that even though the population distribution is skew, as the sample size n increases, the sampling distribution of the mean becomes more and more symmetric and bell-shaped.

What is surprising, though, is that if the sample size is large enough, the sampling distribution of the mean will nearly *always* be this sort of shape, no matter what shape the population distribution is.

The shape of sampling distributions of the mean

For most practical purposes, whatever the shape of the original population distribution, the sampling distributions of the mean for large enough sample sizes are always symmetric and bell-shaped.

These symmetric bell-shaped distributions that we obtain as sampling distributions for large enough values of n are called **normal distributions**.

As you will see in Section 3, these distributions have some very interesting properties which help us to develop the test statistic that we are working towards.

What is a large enough sample?

As a rough guide, you can assume that, whatever the population distribution, for sample sizes greater than 25, the sampling distribution of the mean will always be approximately normal, and in practice, we generally assume that it *is* normal.

In fact, the sampling distribution of the mean will actually be approximately normal for sample sizes (much) smaller than $n = 25$ for many population distributions. On the other hand, there are atypical population distributions for which the sampling distribution of the mean is not (approximately) normal. You will not deal with samples from such populations in M140. This allows us to rephrase a previously highlighted statement.

The shape of sampling distributions of the mean, rephrased

For most practical purposes, whatever the shape of the original population distribution, the sampling distributions of the mean for large enough sample sizes are always *approximately normal*.

Exercises on Section 2

Exercise 4 Means of samples of size 2 from two small populations

Consider the following two small populations of values:

Population A: $10\ 20\ 30\ 40$ and Population B: $10\ 38\ 39\ 40$

(a) Find the sample mean of each of the six different samples of size 2 that you can obtain from Population A. Make a very rough plot of the positions of the six sample means along the horizontal axis.

(b) Repeat what you did in part ((a)) for Population B.

(c) Compare the graphs you obtained in parts ((a)) and ((b)). Which of the two displays a more bell-shaped distribution of sample means? Can you think of a reason why this should be so?

Exercise 5 Change in shape as sample size changes?

The BCS sample with which we are concerned in this unit comprises a total of 679 reading scores (of 7- and 8-year-old children in 2004–2005). We will now *pretend* that this large sample of reading score values is actually the entire population of reading score values. Figure 9 contains pictures of the sampling distributions of the *mean* for samples of various sizes from the (pseudo-)population distribution of reading scores. Describe the changes in shape of these sampling distributions, as the sample size n increases.

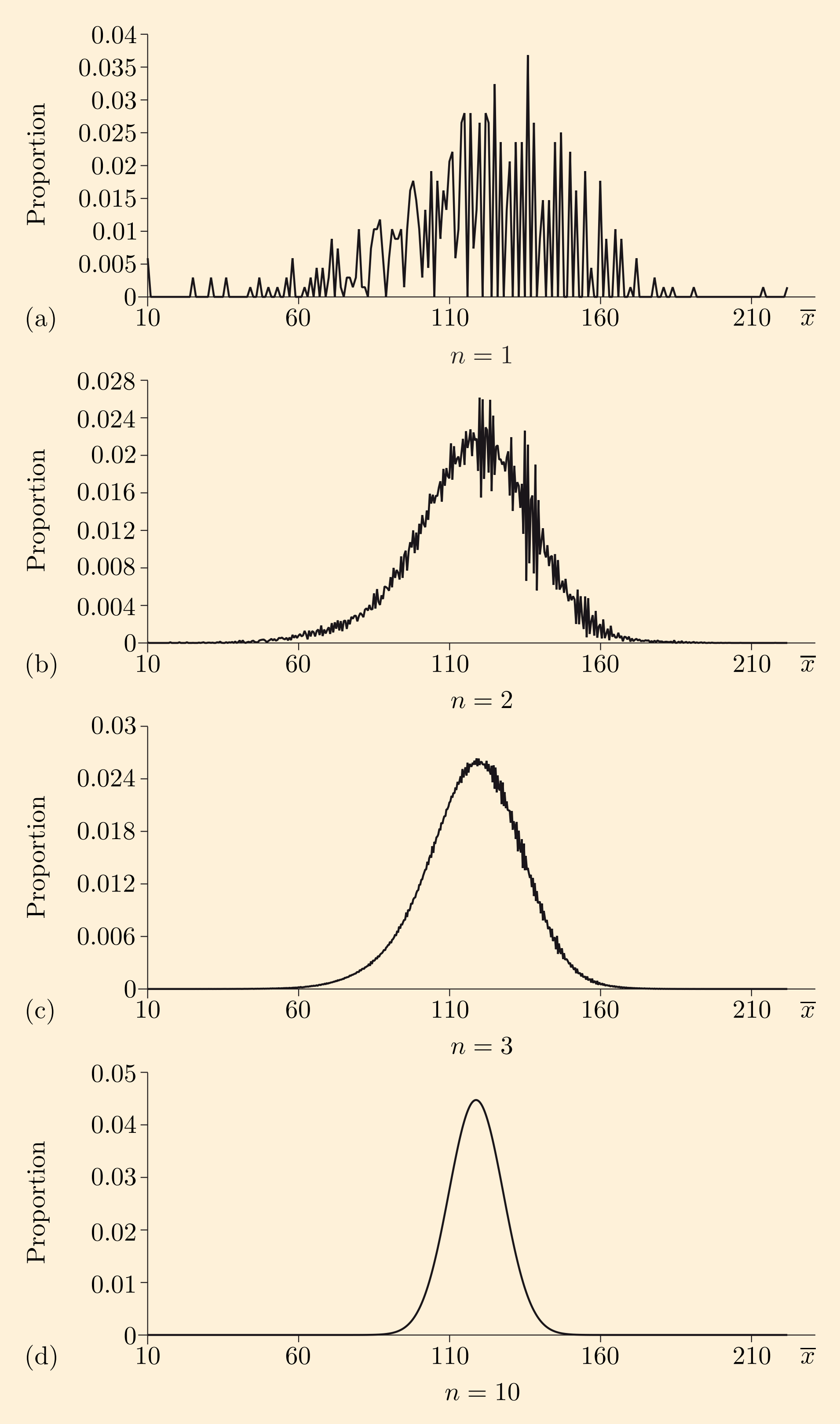


Figure 9 Sampling distributions of the mean as sample size changes

3 Normal distributions

In Section 2 we saw that the sampling distribution of the mean is nearly always approximately normal, provided the sample size is sufficiently large. In this section we examine some of the properties of normal distributions and begin to discover just how important sampling distributions really are.

But first, we need to introduce some important new terminology. You are already familiar with the idea of a sample mean, $\overline{x}$:

$$\overline{x} = \frac{\sum x}{n} = \frac{\text{sum of sample values}}{\text{sample size}}.$$

In this section we shall also need to refer to the **population mean**. For a population of finite – but very large – size, N, this is calculated in exactly the same way, but using *all* the data values in the population. By convention it is labelled μ, so that

The symbol μ is the lower-case Greek letter 'mu', pronounced to rhyme with 'new'.

$$\mu = \frac{\sum x}{N} = \frac{\text{sum of population values}}{\text{population size}}.$$

Because N is often very large indeed, the population is often actually assumed to be of infinite size. For an infinite population, the population mean value, μ, is the mean of a truly enormous sample – the sample size must approach infinity.

There is a similar distinction between the sample standard deviation, s, and the **population standard deviation**, which is denoted by the symbol σ. (This is the lower-case version of the Greek letter 'sigma', which in upper-case form is $\sum$, but there is no connection between the ways that these two symbols are used here.) The formulas are

$$s = \sqrt{\frac{\sum(x-\overline{x})^2}{n-1}} = \sqrt{\frac{\sum x^2 - \left(\sum x\right)^2/n}{n-1}},$$

where the summations are over *sample* values, and

$$\sigma = \sqrt{\frac{\sum(x-\mu)^2}{N}} = \sqrt{\frac{\sum x^2 - \left(\sum x\right)^2/N}{N}},$$

where the summations are over *population* values.

An important property to note is that σ is always a positive number.

Off duty from their work in statistics class μ and σ take a much needed Spring Break in their native Greece

3.1 Normal distributions: location and spread

Normal distributions are important in statistics for two different reasons. You met the first of these in Section 2: many sampling distributions of summary statistics are approximately normal for large enough samples. The other reason is that the distributions of many *populations* are approximately normal. One example is the population of men's heights that you will look at below. In Unit 10 you will see further examples of population distributions that are approximately normal.

Importance of the normal distribution

Normal distributions are important both as (approximate) population distributions, in some cases, and as (approximate) sampling distributions, in many more cases.

The distribution is called the *normal* distribution because it arises so commonly. Normal distributions are also called **Gaussian distributions** after the great German mathematician and scientist C.F. Gauss, who was instrumental in their development. They also appear in popular literature as the 'bell curve'.

Carl Friedrich Gauss (1777–1855)

Carl Friedrich Gauss

Gauss (1777–1855) was a phenomenal mathematician – one of the most productive mathematicians ever. He made exceptional contributions in many fields, perhaps number theory most notably, but also astronomy, geometry, algebra, geophysics and, amongst others, statistics. During the early part of his career he took up the challenge of predicting where Ceres would be found. Ceres was a dwarf planet that had been observed in 1801 but which then disappeared behind the Sun and could not be found when it first reappeared. Gauss developed new methods of estimation and approximation to locate its position. He later published a monograph on the theory of the motion of small planets disturbed by large planets, and in this he introduced several important statistical concepts, including the normal distribution. It is for this reason that the normal distribution is also called the Gaussian distribution, though Gauss did not contribute most to the development of its properties. (The contribution of Laplace (1749–1827), for example, is greater.)

In this unit, we want to explore certain characteristics of normal distributions in order to apply them to sampling distributions. It would be possible to do this exploration using the sort of sampling distributions we met in the last section. However, the descriptions of what is going on tend to look rather complicated, because they involve means of sampling distributions of means. To make things clearer, the exploration is therefore done in the context of a normally distributed population.

Each normal distribution is a precise distribution defined by a mathematical formula involving the mean and standard deviation. We shall not need to use this formula in this module. But despite this mathematical precision, in practice the word 'approximate' is very important above. Real-world populations never have *exact* normal distributions in terms of the mathematical formula; but many are close enough to a normal distribution so that it makes sense to *treat* them as

having normal distributions, in which case we say they are approximately normally distributed.

Figure 10 provides a picture of the population distribution of the heights of all men in Scotland in 2008, based on information given by the *Scottish Health Survey,* 2008.

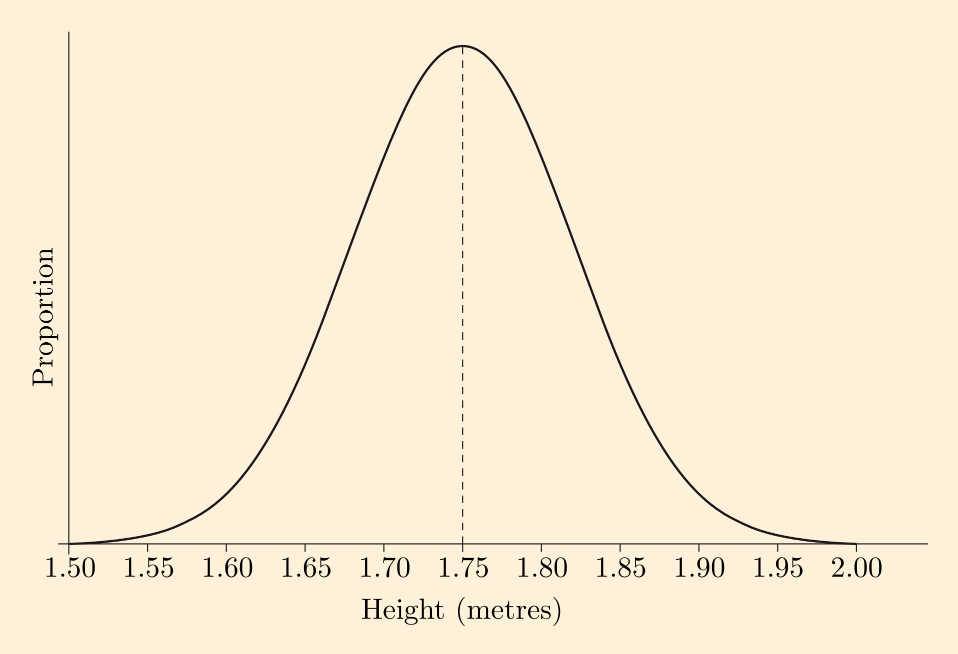


Figure 10 Population distribution of Scottish men's heights (in metres)

This population distribution is very smooth, symmetric and bell-shaped. For the rest of this section, we shall assume that the distribution is indeed normal.

The symmetry of the distribution means that the population mean height is the value corresponding to the mode (peak) of the distribution: about 1.75 metres. (In fact, as well as being the mode and the mean, this value is also the population median!) This characteristic applies more generally so that any normal distribution is symmetric about its mean μ.

Figure 11 shows normal distributions for different values of the mean μ and Figure 12 shows normal distributions for different values of the standard deviation σ.

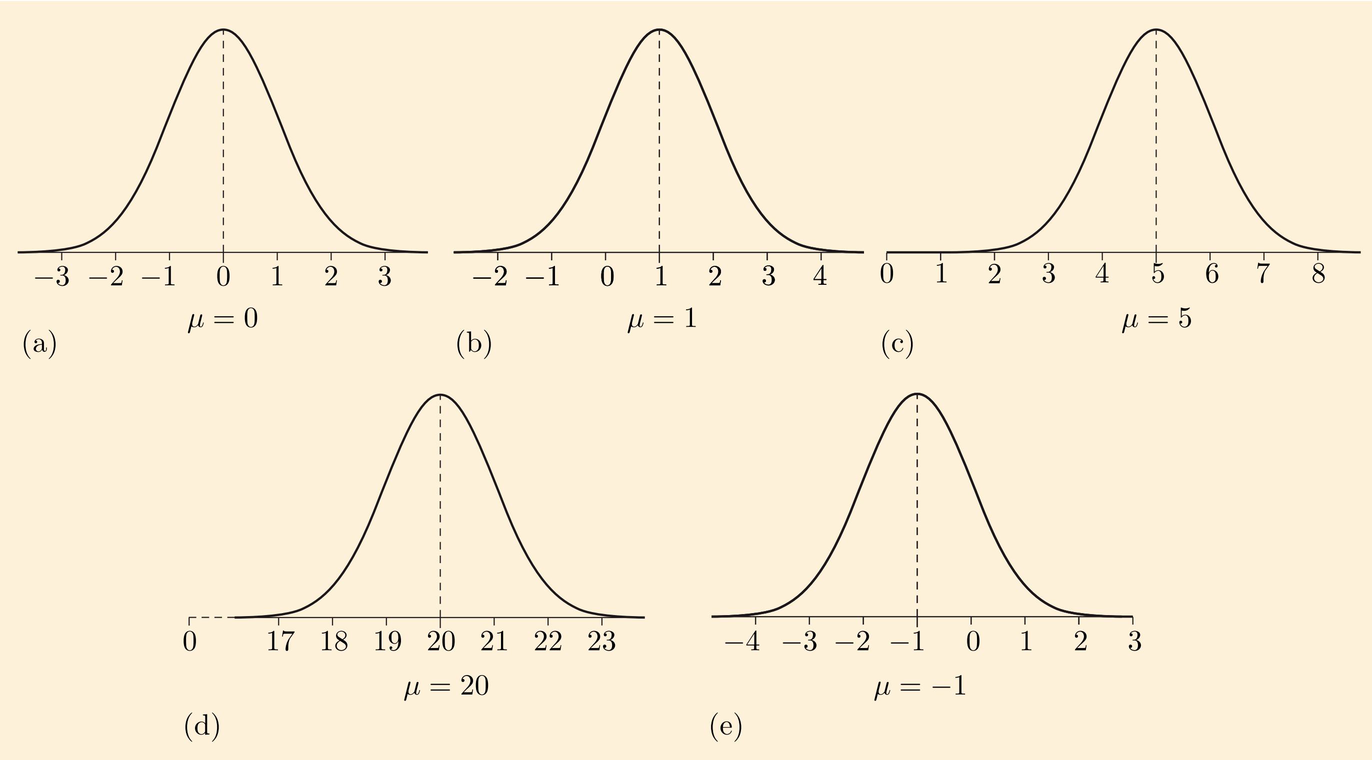


Figure 11 Normal distributions with different locations

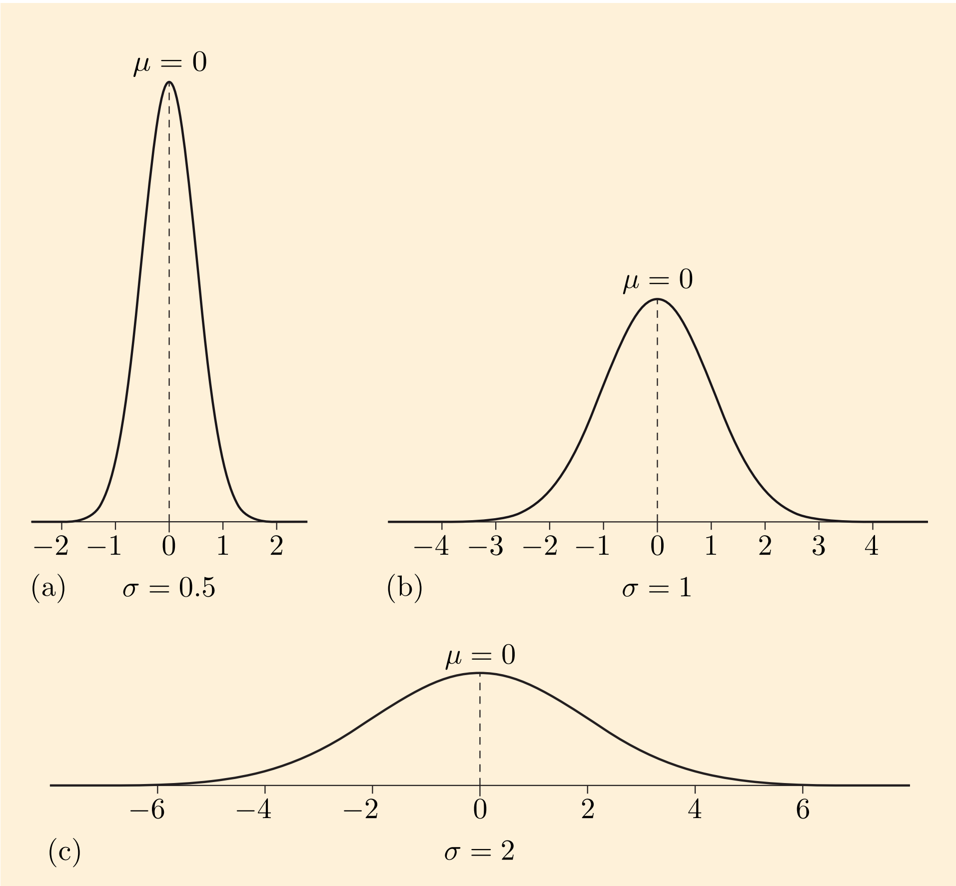


Figure 12 Normal distributions with different spreads

The location of a normal distribution on the horizontal axis depends on the value of its mean μ, as demonstrated by Figure 11.

As with any distribution, the spread of a normal distribution can be measured by the standard deviation of the population, σ. Thus a small value of σ means that the distribution is tightly clustered about the mean; the larger the value of σ, the more spread out the distribution will be – as demonstrated by Figure 12.

Activity 13 What are μ and σ for this normal distribution?

Figure 13 shows another normal distribution. By comparing it with Figures 11 and 12, can you identify the values of μ and σ for this normal distribution?

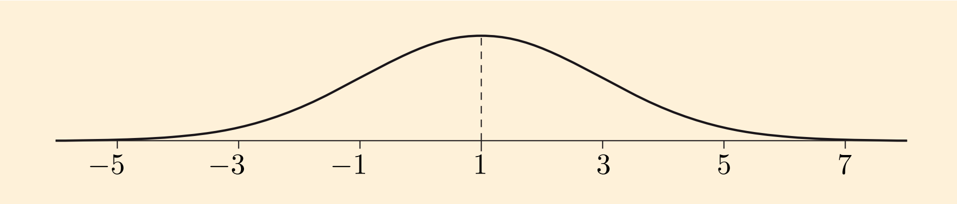


Figure 13 A normal distribution related to those in Figures 11 and 12

Location and spread of the normal distribution: 1

The normal distribution has location specified by the population mean, μ, and spread specified by the population standard deviation, σ.

You have now covered the material needed for Subsection 7.1 of the Computer Book.

You have also now covered the material related to Screencast 1 for Unit 7 (see the M140 website).

3.2 Normal distributions: relating means, standard deviations and plots

For a normal distribution, almost the whole of the distribution (about 99.7%) is contained within plus or minus three standard deviations of the mean. For example, the population distribution of Scottish men's heights (in metres) is normal with mean $\mu \simeq 1.75$ and standard deviation $\sigma \simeq 0.07$. Thus $3\sigma \simeq 0.21$, and so almost the whole of the distribution is contained within plus or minus 0.21 metres of the mean 1.75 metres (i.e. between $1.75 - 0.21 = 1.54$ metres and $1.75 + 0.21 = 1.96$ metres). You can check for yourself in Figure 14 – which is an annotated copy of Figure 10 – that this is indeed the case.

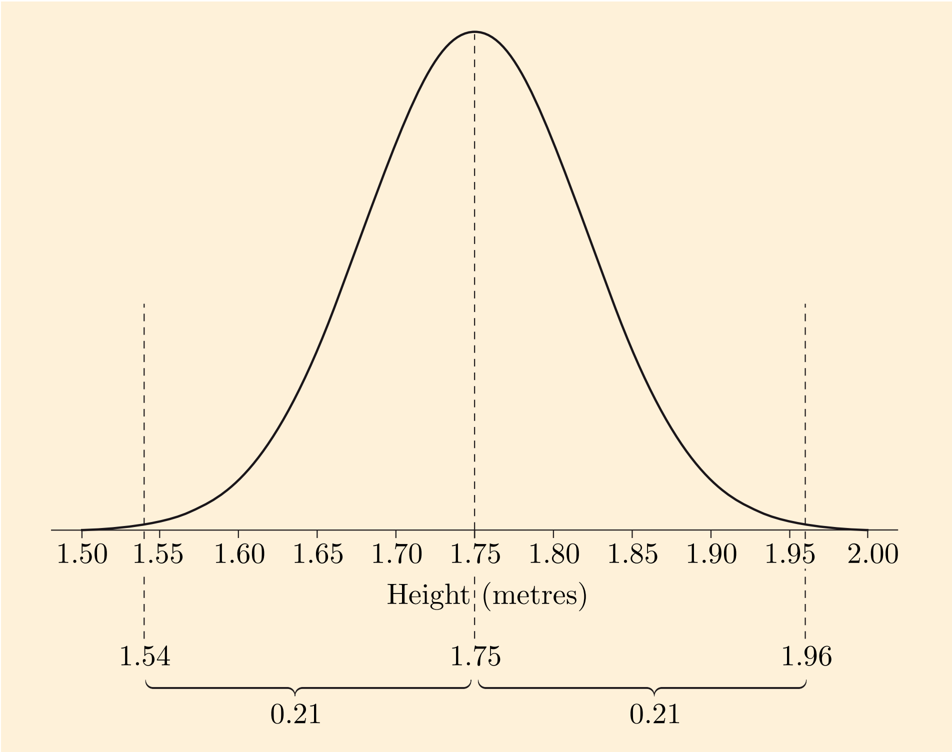


Figure 14 Annotated population distribution of Scottish men's heights

(Similar percentages are known for all other numbers of standard deviations; for instance, 95.4% of the distribution is contained within plus or minus two standard deviations, and 68.3% within plus or minus one standard deviation.)

Location and spread of the normal distribution: 2

The normal distribution has its mode at μ, and almost the whole of the normal distribution is contained between $\mu - 3\sigma$ and $\mu + 3\sigma$.

The links between the graph of a normal distribution and its mean and standard deviation suggest that a picture of the distribution can be used to obtain approximate values for its mean and standard deviation.

Example 2 Approximate values for the mean and standard deviation

The population distribution of a certain variable x is known to be normal. This distribution is pictured in Figure 15.

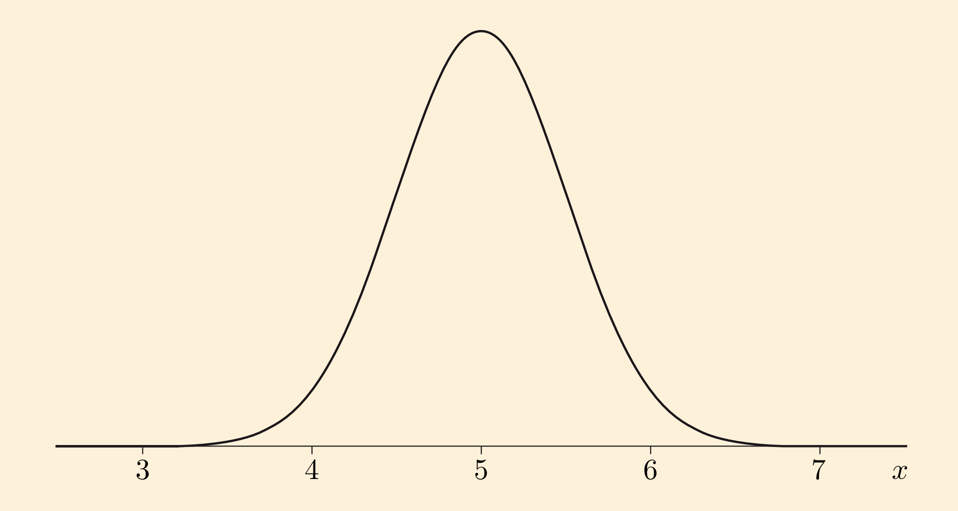


Figure 15 A normal distribution

The mode of this normal distribution occurs at about $x = 5$. This means that the population mean must be approximately equal to 5. So, $\mu \simeq 5$. We say *approximately* equal because μ may not be exactly equal to 5. It could be 5.1 or 4.9; it is impossible to give an exact value here.

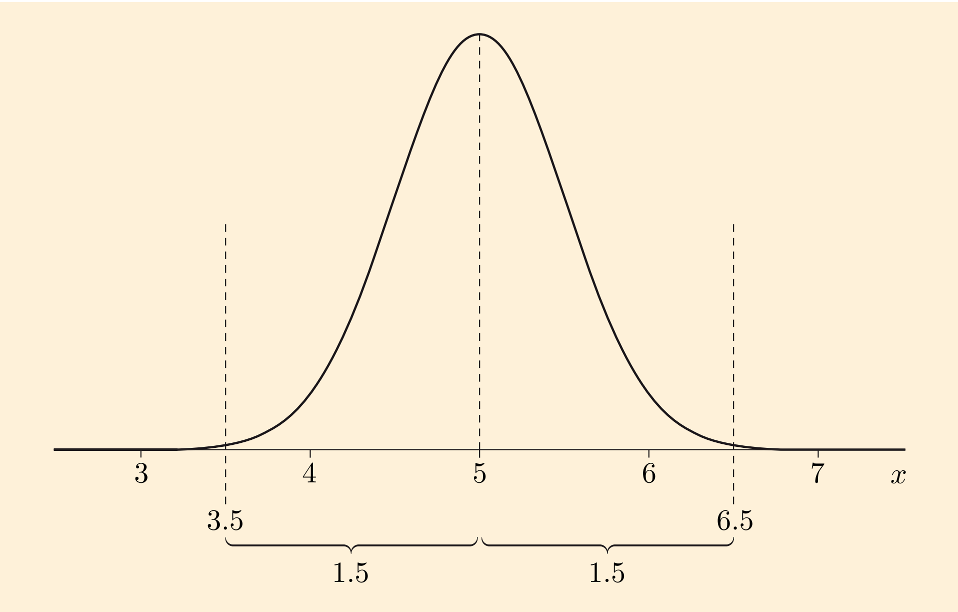


Figure 16 Investigating the spread of a normal distribution

The dashed lines in Figure 16 indicate that almost all of the distribution is contained between $x = 3.5$ and $x = 6.5$ (i.e. within 5 ± 1.5). This means that $3\sigma \simeq 1.5$, so $\sigma \simeq 0.5$.

In summary, the normal distribution plotted in Figure 15 is approximately the normal distribution with mean $\mu = 5$ and standard deviation $\sigma = 0.5$.

Activity 14 Approximate values for the mean and standard deviation

For each of the normal distributions shown in the parts of this activity, find approximate values for the mean and standard deviation, using the method described above.

(a)

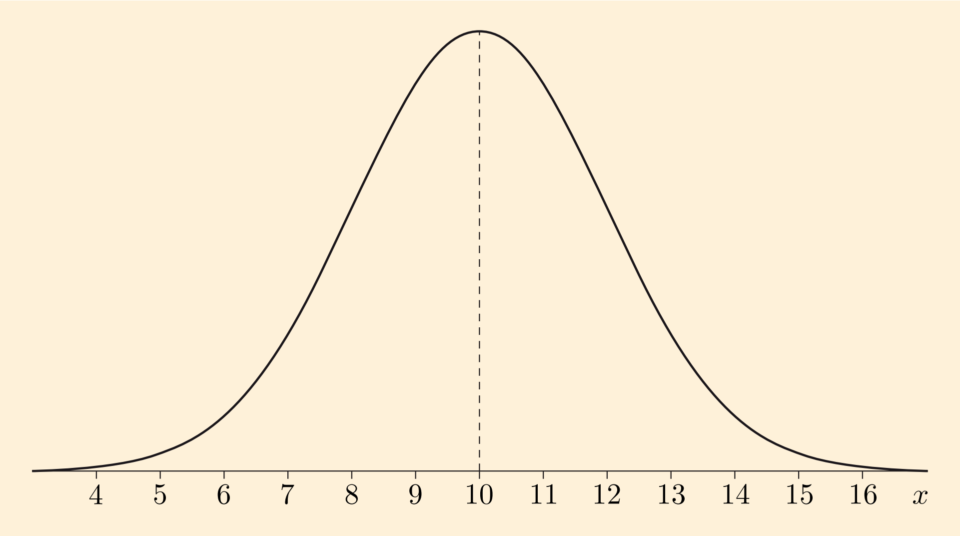


Figure 17 Another normal distribution

(b)

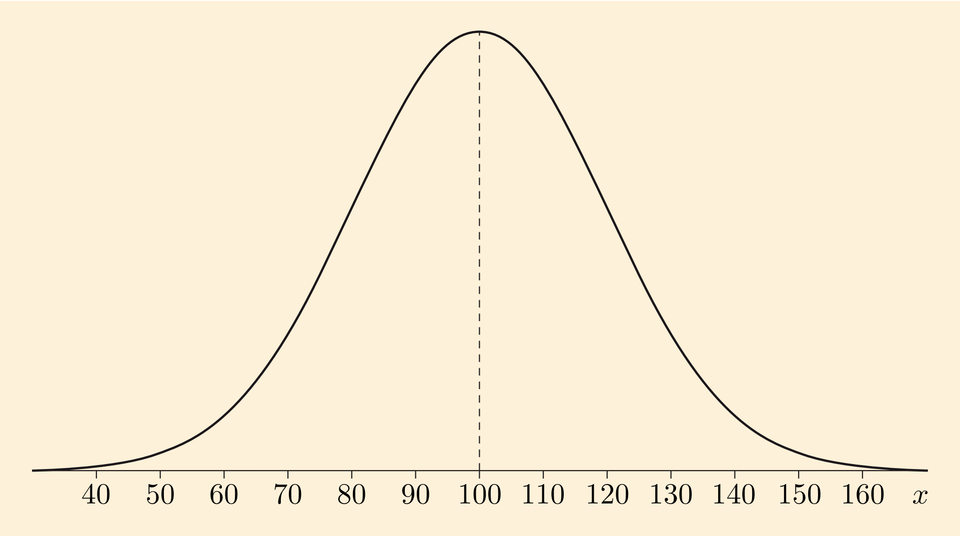


Figure 18 Yet another normal distribution

(c)

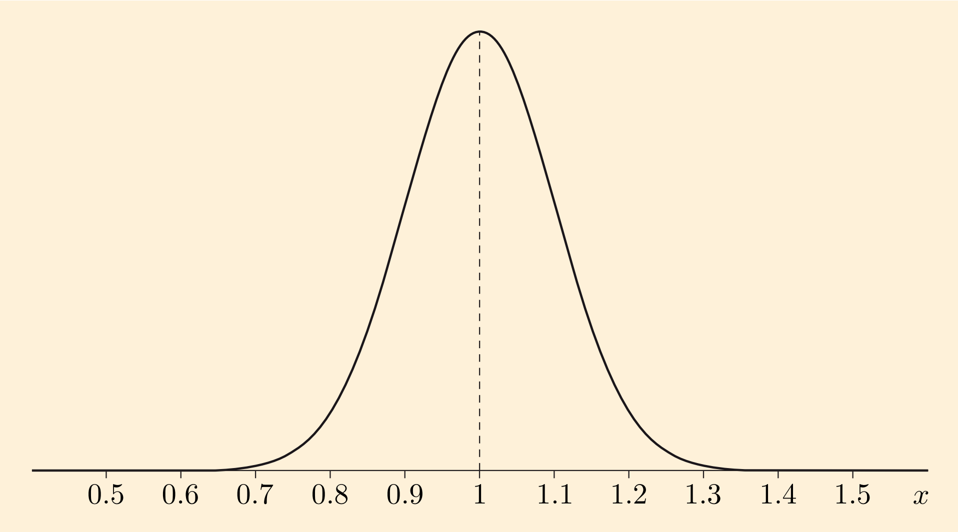
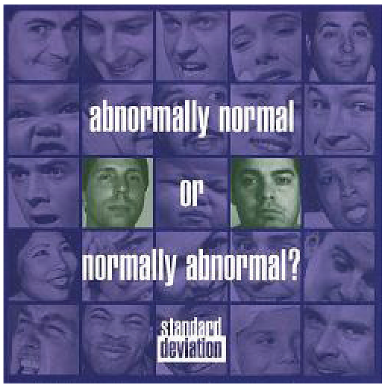


Figure 19 And yet one more normal distribution

Conversely, knowing the mean and standard deviation of a normal distribution enables us to make a rough sketch of the distribution. Any sketch of a normal distribution will show a symmetric and bell-shaped curve. More specifically, the distribution must be symmetric about the mean. In addition, almost the whole of the distribution must be contained within plus or minus three standard deviations of the mean.



The 1997 music CD ‘abnormally normal or normally abnormal?’ by the band ‘standard deviation’.

Example 3 Sketching a normal distribution

The normal distribution of a variable x has mean $\mu = 15$ and standard deviation $\sigma = 3$. To sketch this distribution, draw a symmetric, bell-shaped curve centred on the value of μ, which in this case is 15. The standard deviation is $\sigma = 3$, so that $3\sigma = 9$. We therefore know that just about all the distribution is contained within 15 ± 9 (i.e. lies between $15 - 9 = 6$ and $15 + 9 = 24$). A sketch of the distribution can therefore be drawn and should resemble Figure 20.

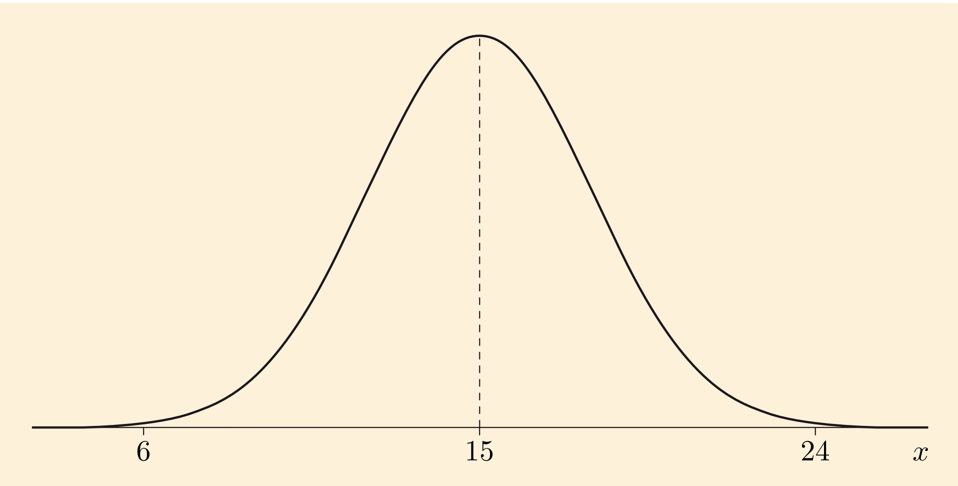


Figure 20 The normal distribution with $\mu = 15$ and $\sigma = 3$

The scale that is used for the horizontal axis certainly affects the shape of the normal distribution, as demonstrated by Figure 21. The important thing, though, is that the information conveyed by the sketch remains exactly the same.

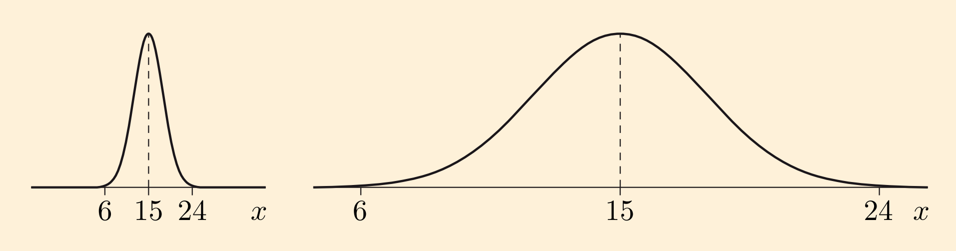


Figure 21 The same normal distribution plotted on different horizontal scales

Also, for the aspects we are investigating, the height of the distribution does not really matter; all the information we require about the relationship between the distribution and its mean and standard deviation is provided by the scale on the horizontal axis. For this reason there is no need to bother with a vertical scale at all.

Activity 15 Sketching a normal distribution

Sketch the following distributions:

- The normal distribution of a variable x with mean 1000 and standard deviation 100.
- The normal distribution of a variable x with mean 2 and standard deviation 0.25.

Activity 15 demonstrates that it always makes sense to think of the horizontal axis of a normal distribution in terms of the number of standard deviations of the variable away from the mean. This is illustrated in Figure 22, which is an important picture in understanding the normal distribution. Notice how the horizontal scale is marked off using μ and σ.

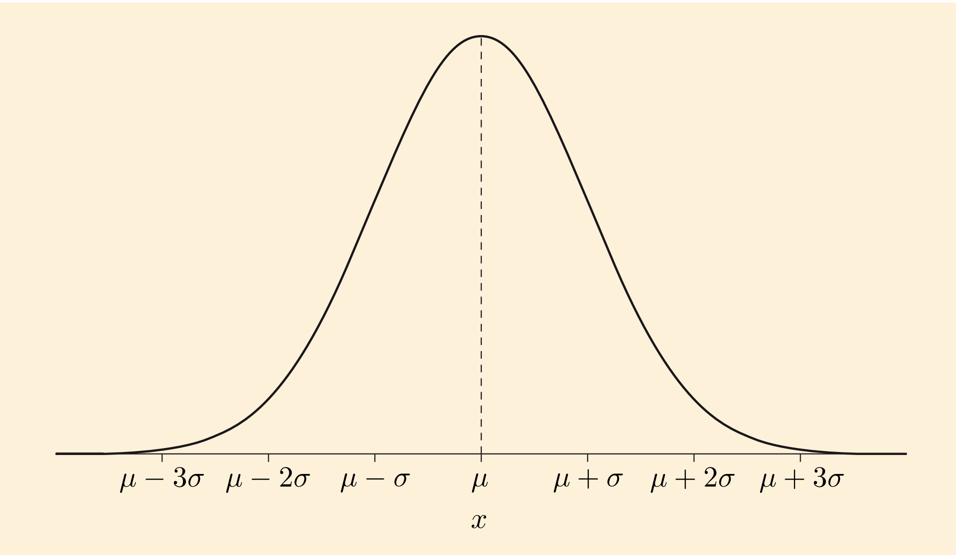


Figure 22 The normal distribution with its scale marked in terms of μ and σ

You have now covered the material related to Screencast 2 for Unit 7 (see the M140 website).

3.3 The standard normal distribution

We can go one step further than that represented by Figure 22 (Subsection 3.2) and think of all normal distributions in terms of one special normal distribution. This special normal distribution has mean zero and standard deviation one, and is called the **standard normal distribution**. It looks like Figure 23. Figure 23, in turn, looks like Figure 22 with μ and σ in the labels on the horizontal axis replaced by 0 and 1, respectively: so, $\mu - 3\sigma$ has become $0 - (3 \times 1) = -3$, $\mu - 2\sigma$ has become $0 - (2 \times 1) = -2$, and so on.

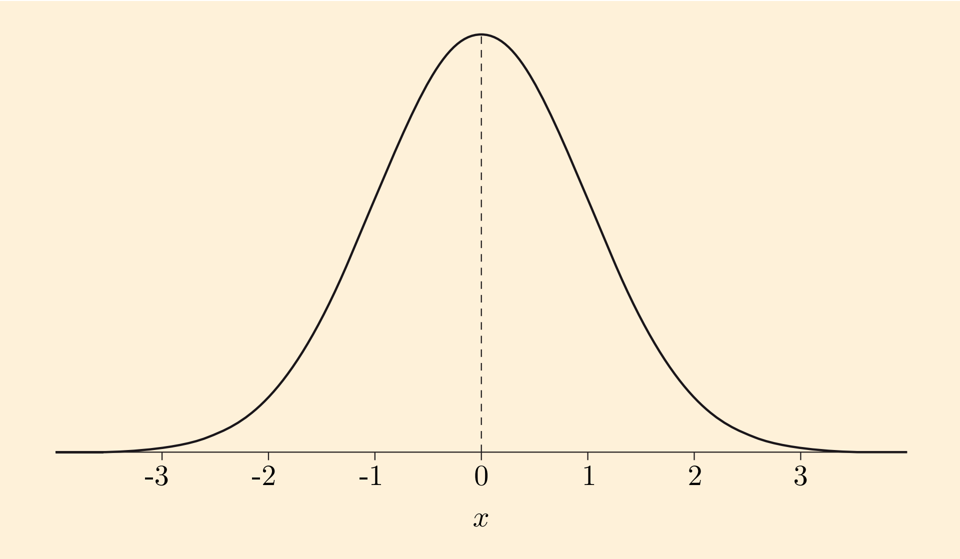


Figure 23 The standard normal distribution

The standard normal distribution

The standard normal distribution is the particular normal distribution that has mean $\mu = 0$ and standard deviation $\sigma = 1$.

It turns out that we can *transform* all normal distributions to the standard normal distribution.

Example 4 Transforming to the standard normal distribution

The normal distribution of a variable x with mean $\mu = 10$ and standard deviation $\sigma = 2$ is illustrated in Figure 24.

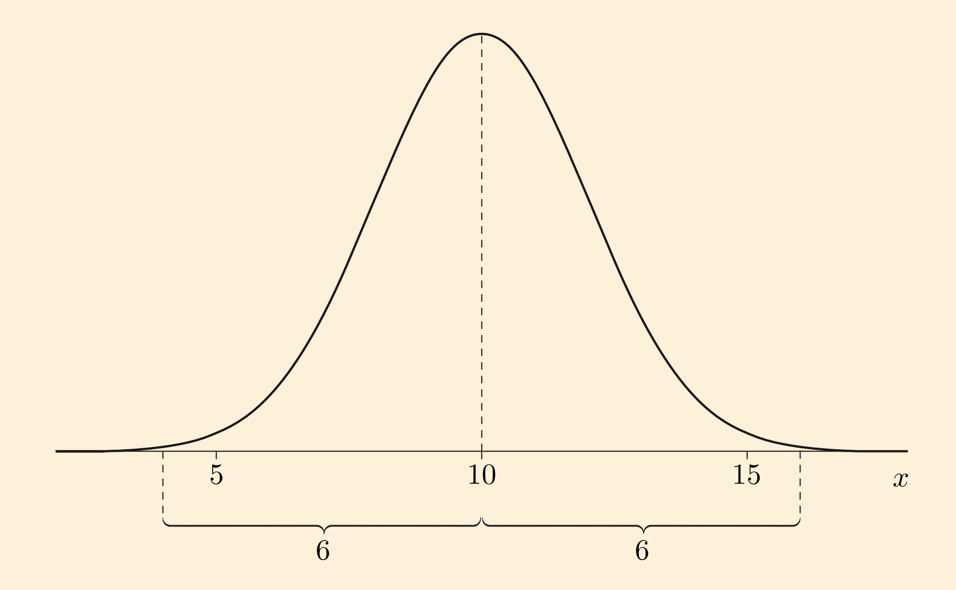


Figure 24 The normal distribution with $\mu = 10$ and $\sigma = 2$

First, we can shift the whole of the distribution to the left so that the mode occurs at zero just by subtracting 10 from each value of x. This is shown in Figure 25. It changes the *location* of the distribution but leaves the *spread* unchanged.

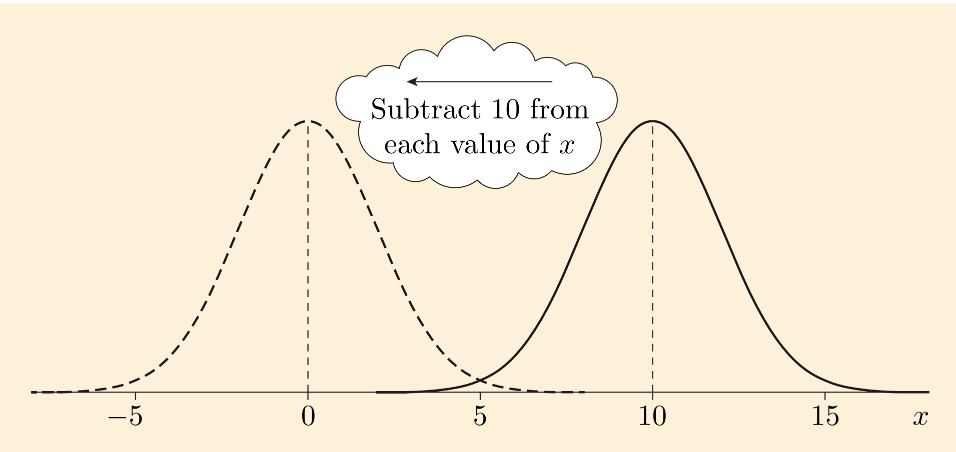


Figure 25 Shifting the distribution of x

The dashed curve in Figure 25 is now a new normal distribution with mean zero and standard deviation 2. This new distribution is the distribution of the variable v, say, where $v = x - 10$. The normal distribution of v differs from the standard normal distribution only by having standard deviation 2 rather than 1. However, if we now think of the horizontal axis in terms of the number of standard deviations of v away from the mean, then we obtain Figure 26.

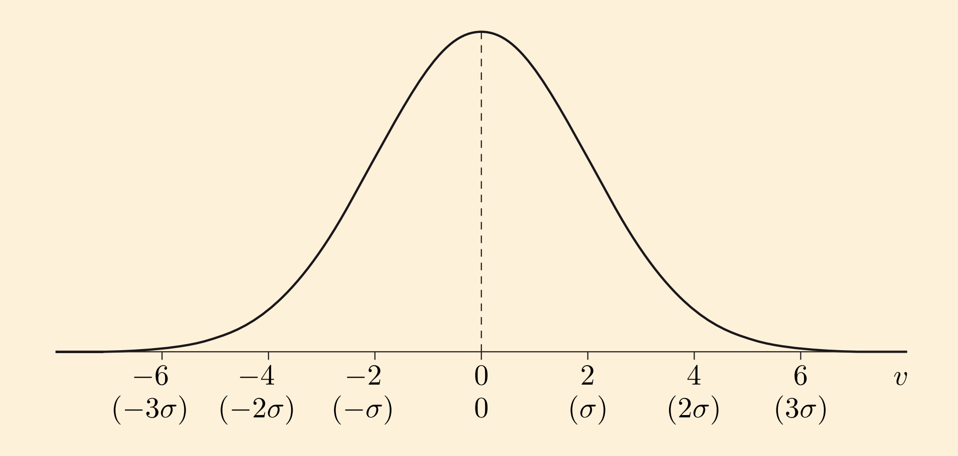


Figure 26 The normal distribution of $v = x - 10$ with mean 0 and standard deviation 2

Then, dividing every value of v by the standard deviation 2 gives the distribution of $v/2$. This distribution, shown in Figure 27, is the standard normal distribution, as required.

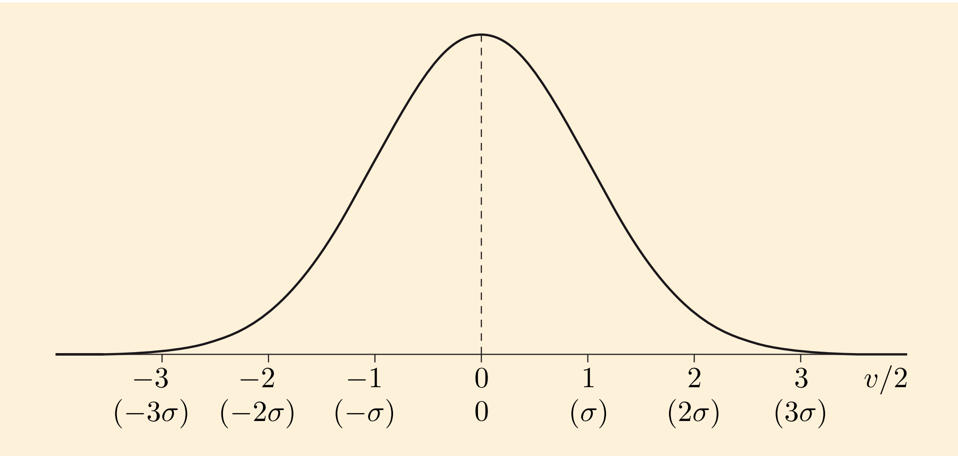


Figure 27 The normal distribution of $v/2$ with mean 0 and standard deviation 1

We have shown that if the variable x has a normal distribution with mean 10 and standard deviation 2, then the variable $v/2 = (x - 10)/2$ has the standard normal distribution.

Example 4 is a specific example of the following general result. If we start with a normal distribution for x, with mean μ and standard deviation σ, then:

- By subtracting μ from each value of x we obtain the distribution of $v = x - \mu$. This distribution is normal with mean zero and standard deviation σ.
- By then dividing each value of v by σ we obtain the variable $z = v/\sigma$, which has the standard normal distribution.

Combining the formulas for z and v, we find that

$$z = \frac{x - \mu}{\sigma}.$$

Transforming a normal distribution to the standard normal distribution

If a variable x has a normal distribution with mean μ and standard deviation σ, then the variable

$$z = \frac{x - \mu}{\sigma}$$

has the standard normal distribution.

Activity 16 Transforming some particular normal distributions

For the normal distributions with the following values of μ and σ, write down the appropriate formula to transform the variable x to the variable z that follows the standard normal distribution.

(a) $\mu = 10, \sigma = 2$ (b) $\mu = 100, \sigma = 20$ (c) $\mu = 1, \sigma = 0.1$

Activity 17 Transforming the distribution of Scottish men's heights

(a) Assume that the population distribution of Scottish men's heights h (in metres) is normal with mean $\mu = 1.75$ and standard deviation $\sigma = 0.07$. Write down the formula for z which transforms each value of the variable h to the number of standard deviations from its mean.

(b) Calculate the value of z corresponding to each of the following values of h (in metres). In each case, interpret your answer by completing a sentence of the form 'So a height of *** metres is *** standard deviations *** the mean height of *** metres'.

$$h = 1.96; \quad h = 1.61; \quad h = 1.785.$$

Importance of the standard normal distribution

The development in this subsection implies that by describing every normal distribution in terms of z, the number of standard deviations by which the variable differs from its mean, we can think of all normal distributions in terms of just one distribution: the standard normal distribution.

Wall space increased at the Statistics Art Gallery when it became clear that only one picture was required in the 'normal distribution' collection.

You have now covered the material needed for Subsection 7.2 of the Computer Book.

You have also now covered the material related to Screencast 3 for Unit 7 (see the M140 website).

Exercises on Section 3

Exercise 6 Approximating the mean and standard deviation

Find approximate values for the mean and standard deviation of the normal distribution shown below.

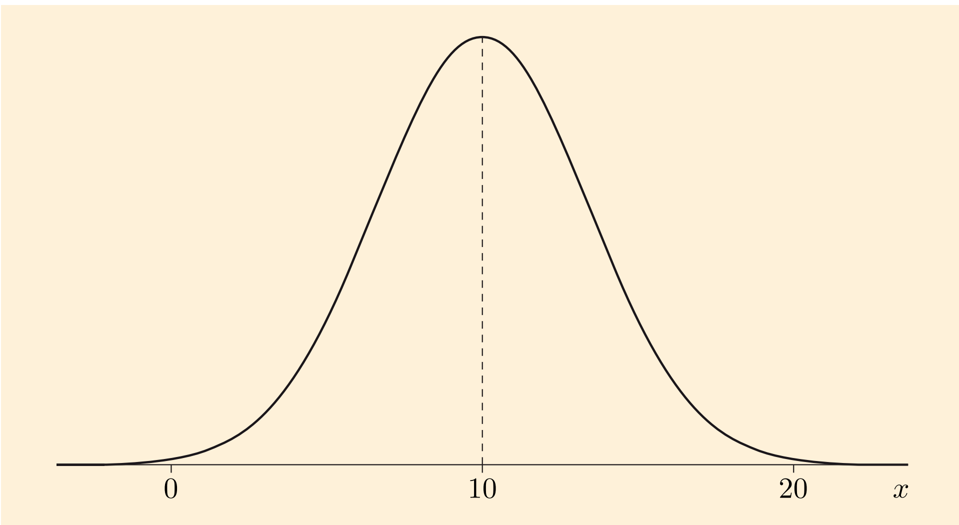


Figure 28 Yet again, a normal distribution

Exercise 7 Approximating another mean and standard deviation

Find approximate values for the mean and standard deviation of the normal distribution shown below.

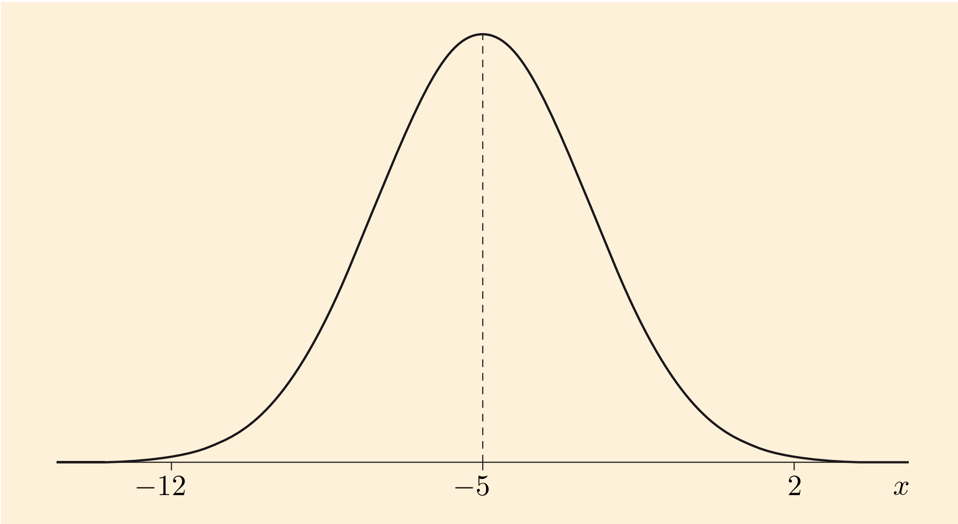


Figure 29 And one more time, another normal distribution

Exercise 8 Sketching a normal distribution

The normal distribution of a variable x has mean -1 and standard deviation 1. Sketch the distribution.

Exercise 9 Sketching another normal distribution

The normal distribution of a variable x has mean 4 and standard deviation 4. Sketch the distribution.

Exercise 10 Obtaining z for a normal distribution

Write down the appropriate formula to transform the variable x to the variable z that follows the standard normal distribution when

(a) x has the normal distribution with mean 6 and standard deviation 3.3;

(b) x has the normal distribution with mean -6 and standard deviation 2.

Exercise 11 Calculating z from x

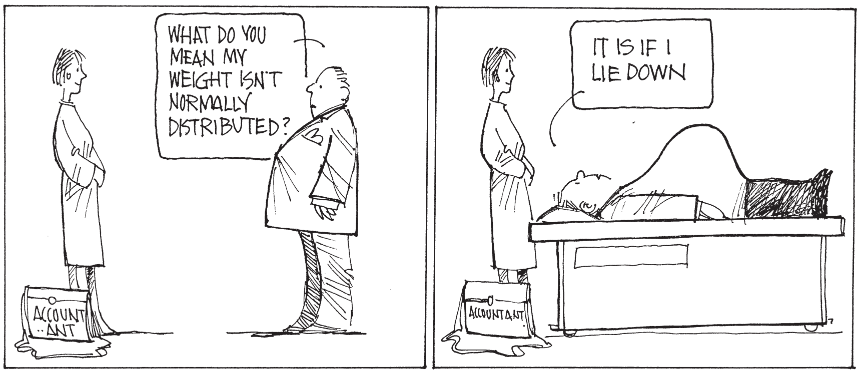
Assume that x follows the normal distribution with mean $\mu = 2$ and standard deviation $\sigma = 10$. Write down the appropriate formula for z which transforms the variable x to the number of standard deviations from its mean. Calculate the value of z corresponding to $x = 3$.

Exercise 12 Calculating z from x for another normal distribution

Assume that x follows the normal distribution with mean $\mu = -1$ and standard deviation $\sigma = 0.5$. Write down the appropriate formula for z which transforms the variable x to the number of standard deviations from its mean. Calculate the value of z corresponding to $x = 0$.

4 Sampling distributions re-revisited

We now take a closer look at the sampling distributions of the sample mean that you met in Section 2. As we said there, provided the sample size is sufficiently large (roughly speaking, greater than 25), these sampling distributions are approximately normal. Thus the ideas discussed in Section 3, which apply to *all* normal distributions, apply (approximately) to these sampling distributions as well. These ideas will enable us to find a suitable test statistic to use for testing some of the hypotheses we are interested in for the BCS survey.



We begin by examining the relationship between sampling distributions of the mean and the original population distribution in a little more detail.

Activity 18 Means of distributions of sample means

(a) Consider again the population distribution of MS221 examination marks which you met in Section 2. In fact, this population distribution has mean $\mu = 66$ and standard deviation $\sigma = 22$. Figure 30 shows the sampling distributions of the mean for various sample sizes. (Figure 30 is similar to Figure 6 but for some different values of n.)

What do you notice about the means of these sampling distributions compared with the population mean?

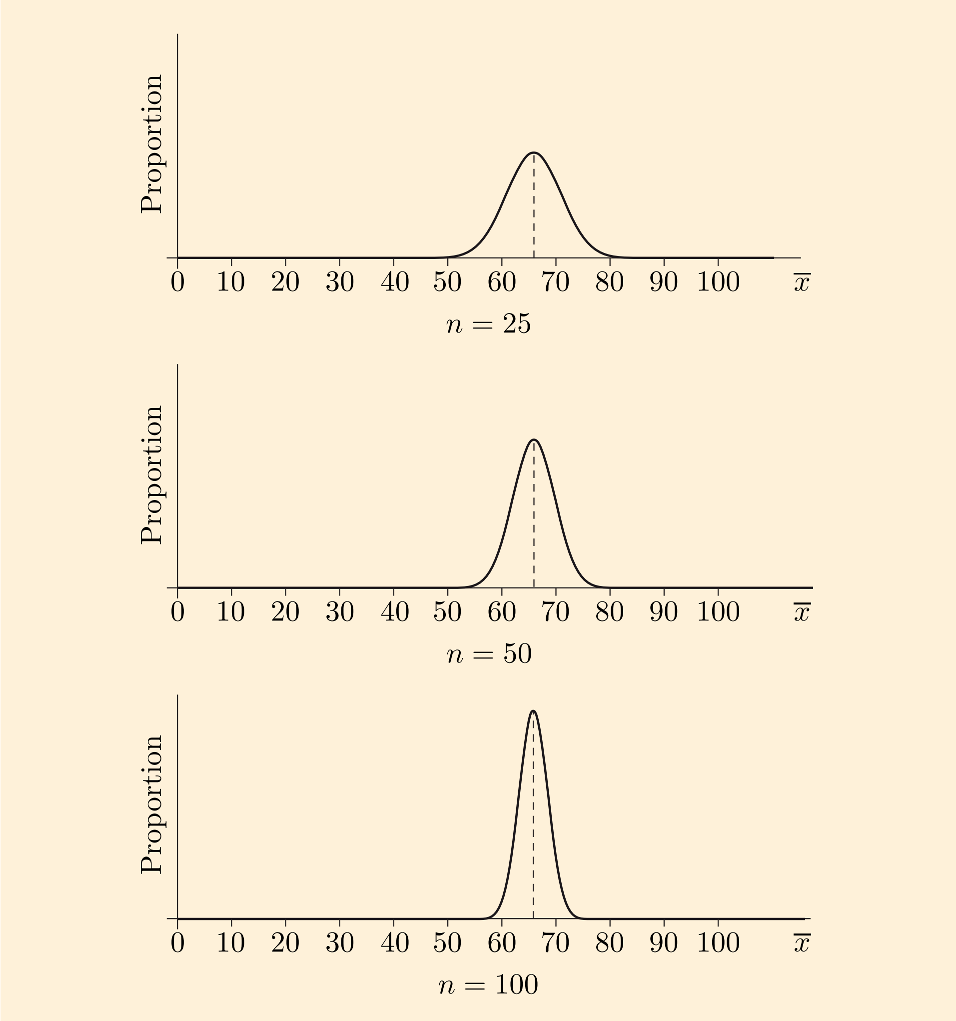


Figure 30 Sampling distributions of the mean for samples of size n from the population of exam marks

(b) Consider again the population distribution of full-time employees' earnings which you met in Example 1, in Section 2. This population distribution has mean $\mu = 491$ and standard deviation $\sigma = 283$ (in \$). Figure 31 shows again the sampling distributions of the mean for various sample sizes. (Figure 31 is similar to Figure 8 but for different values of n.)

What do you notice about the means of these sampling distributions compared with the population mean?

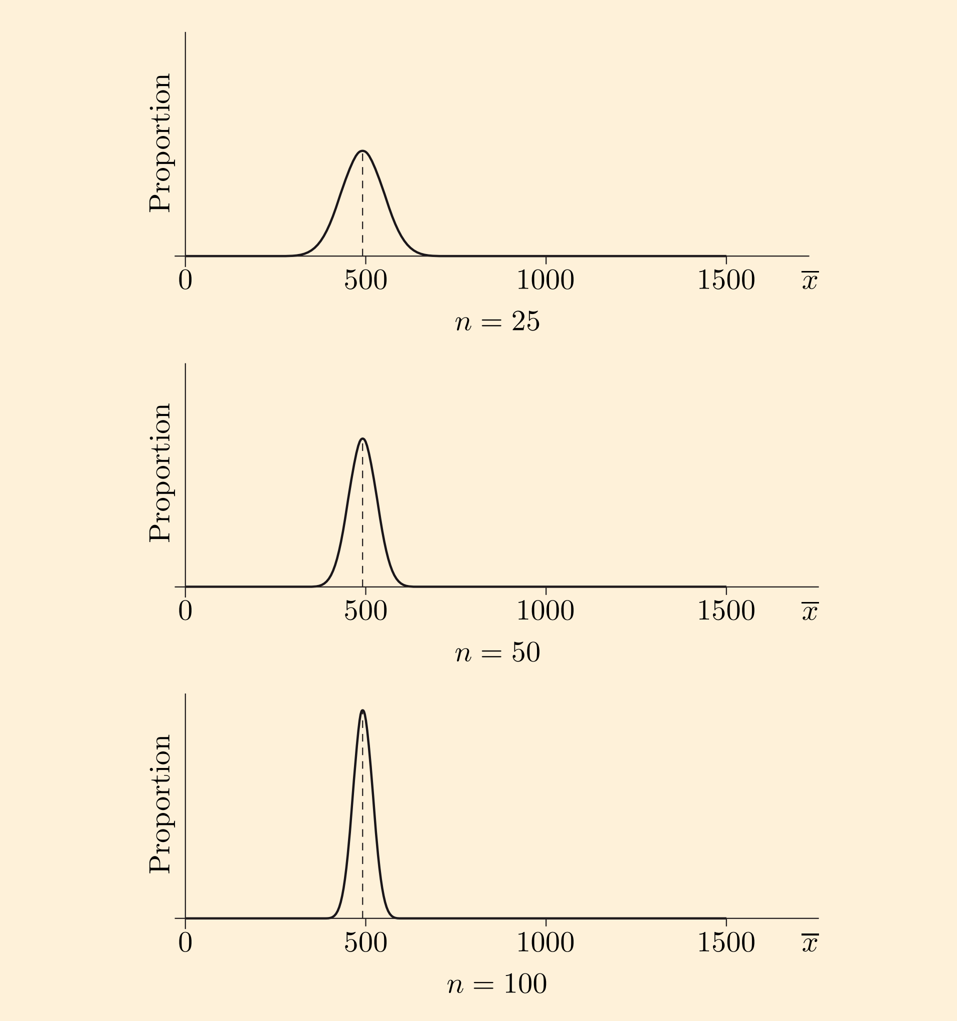


Figure 31 Sampling distributions of the mean for samples of various sizes from the population of employees' earnings

The conclusions of Activity 18 hold more generally so that *whatever* the population distribution (no matter what shape) and *whatever* the sample size (no matter how small), the mean of the sampling distribution is always equal to the population mean μ.

Now let us take a closer look at the *spread* of the sampling distributions.

Activity 19 Standard deviations of distributions of sample means: 1

(a) Consider again the sampling distributions of the mean for MS221 examination marks that are shown in Figure 30. What do you notice about the standard deviations of these sampling distributions?

(b) Consider again the sampling distributions of the mean for full-time employees' earnings that are shown in Figure 31. What do you notice about the standard deviations of these sampling distributions?

In fact it can be shown that for population standard deviation σ, the standard deviation of the sampling distribution of the mean for samples of size n is $\sigma/\sqrt{n}$.

Activity 20 Standard deviations of distributions of sample means: 2

The population distribution of examination marks has standard deviation $\sigma = 22$. Use the formula to find the standard deviation of the sampling distribution of the mean for samples of size

(a) 25; (b) 50; (c) 100.

Both the formula $\sigma/\sqrt{n}$ and the calculations in Activity 20 confirm that the standard deviation of the sampling distribution of the mean does decrease as n increases, as was suggested in Activity 19. What is not so clear, and is perhaps unexpected, is the precise way in which the standard deviation of the sampling distribution of the mean depends on n – through its square root.

The terminology 'standard *error*' is related to the notion of 'sampling *error*', which you met in Subsection 4.1 of Unit 4.

The expression 'standard deviation of the sampling distribution of the mean' is a bit of a mouthful. It is often referred to as the **standard error of the mean** for samples of size n, or sometimes just the **standard error** for short, in which case it can be abbreviated to the symbol SE. Using this abbreviation, we obtain the formula $\text{SE} = \sigma/\sqrt{n}$, which is easier to remember.

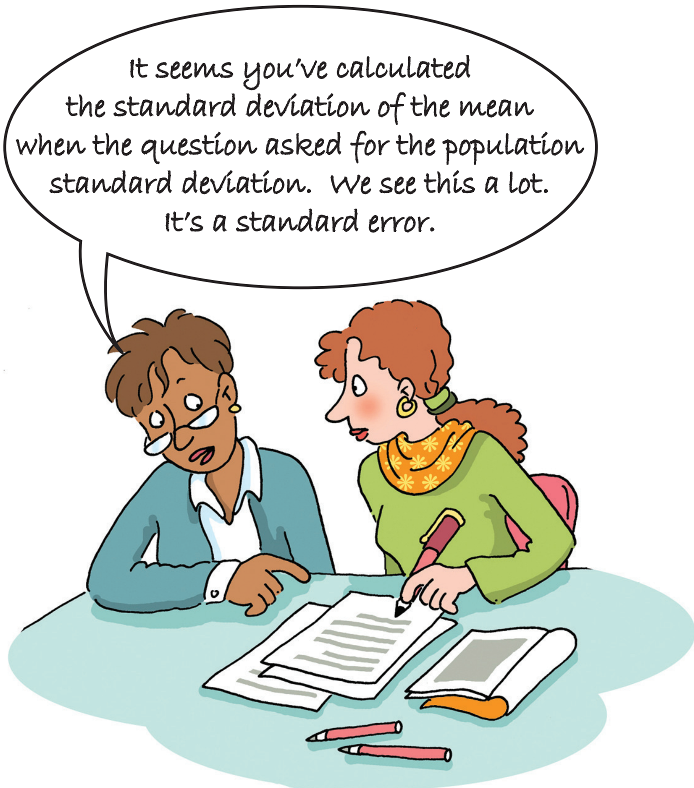
The above result holds generally for *all* sampling distributions, no matter *what* the population distribution and no matter *what* sample size is involved. So there is a very precise relationship between sampling distributions and the population distribution. It can be summarised as follows.

Mean and standard deviation of the sampling distribution of the mean

- The mean of the sampling distribution is equal to μ, the population mean.
- The standard deviation of the sampling distribution is called the standard error of the mean. It is given by

$$\text{SE} = \frac{\sigma}{\sqrt{n}},$$

where n is the sample size and σ is the population standard deviation.



The relationship between sampling distributions and the population distribution is particularly useful when the sample size is large and the sampling distribution is approximately normal. In practice, we usually have very little information about the population distribution itself. Indeed we often have only a sample of data on which to base our analysis; there is *no* other information about the population. Yet many techniques of statistical inference require us to make some assumptions about the population distribution.

The advantage of working with large samples is that, no matter what shape the population distribution is, the sampling distribution of the mean for samples of size n will *always* be more or less normal. Moreover, we know that the mean of this sampling distribution is equal to the population mean, μ, and the standard deviation is the standard error, given by SE $= \sigma/\sqrt{n}$, where σ is the population standard deviation. This is summarised below.

Approximate normality of the sampling distribution of the mean

If n is large, no matter what shape the population distribution is, the sampling distribution of the mean for samples of size n will be approximately normal with mean equal to the population mean, μ, and standard deviation equal to the standard error, SE $= \sigma/\sqrt{n}$.

(This important result is often called the *central limit theorem*.)

Activity 21 Approximate distribution of ball bearing diameters

The population distribution of the diameters of ball bearings produced by a particular manufacturer has mean $\mu = 2$ mm and standard deviation $\sigma = 0.01$ mm. Find the standard deviation of the sampling distribution of the mean for samples of 25 such ball bearings. Hence give the approximate

distribution of the mean diameter of ball bearings in samples of size 25.

What this implies is that we can base our analysis on the relationship between the sample data and the sampling distribution of the mean. Thus we infer back from the evidence provided by the sample data to the sampling distribution. Then our knowledge of the links between this sampling distribution and the population distribution allows us to draw conclusions about the population mean. This is a very important strategy in statistics.

The new hypothesis test, the z-test, is based on just this principle and will be fully discussed in Section 5. As you now know, the sampling distribution of the mean, $\overline{x}$, for large samples of size n is approximately normal with mean μ and standard deviation SE $= \sigma/\sqrt{n}$. As with any normal distribution, we can transform this normal sampling distribution into the standard normal distribution. This means that the distribution of the variable

$$z = \frac{\overline{x} - \mu}{\text{SE}}$$

is the standard normal distribution (with mean zero and standard deviation one). There is a strong connection between this result and the z-test to follow.

You have now covered the material related to Screencast 4 for Unit 7 (see the M140 website).

Exercises on Section 4

Exercise 13 Standard deviations of the mean as sample size changes

The population distribution of full-time employees' earnings has standard deviation $\sigma = 283$. Find the standard deviation of the sampling distribution of the mean for samples of size

(a) 9; (b) 25; (c) 100.

Exercise 14 Standard deviations of another mean

The population distribution of a certain quantity has standard deviation $\sigma = 3.6$. Find the standard deviation of the sampling distribution of the mean for samples of size

(a) 4; (b) 19; (c) 300.

Exercise 15 Standard deviation of the average content of water bottles

The population distribution of the amount of water contained in a nominally one-litre bottle from a certain manufacturer has mean $\mu = 1.01$ litres and standard deviation $\sigma = 0.01$ litres. Find the standard deviation of the sampling distribution of the mean for samples of 40 such bottles. Hence give the approximate distribution of the mean amount of water contained in samples of 40 one-litre bottles from this manufacturer.

5 The one-sample z-test

In this section we shall develop a new hypothesis test, the **one-sample z-test**. The hypotheses are concerned with the mean, μ, of the population from which the sample is selected. We shall suppose that a particular value, say A, is of special interest as a potential value for μ. The null hypothesis is

Unless we need to distinguish the one-sample z-test from the two-sample z-test that will be developed in Section 6, we often omit the phrase 'one-sample'.

$$H_0\colon \mu = A,$$

and the alternative hypothesis is

$$H_1\colon \mu \neq A.$$

Alternative hypotheses of this form are often called **two-sided alternative hypotheses**. This is because they include both $\mu < A$ and $\mu > A$.

One-sided alternative hypotheses will be discussed in Unit 10.

The above is the first of the four stages of hypothesis testing that you were introduced to at the start of Section 4 of Unit 6. In abbreviated form, these are:

(a) Set up the hypotheses that we wish to test.

(b) Determine the sampling distribution of a test statistic under the assumption that the null hypothesis is true.

(c) Ascertain how unlikely the observed value of the test statistic is on the basis of the sampling distribution.

(d) If the test statistic turns out to have a very unlikely value, then either:
- a very unusual event has happened, or
- the sample has provided evidence against the correctness of the null hypothesis.

To develop ideas in the current context, we first consider the simpler case where the population standard deviation is assumed to be known, and in Subsection 5.2 we consider the more realistic case where it is unknown. The tests that are developed make use of the results presented in Section 4 about the sampling distribution of the sample mean.

5.1 The z-test with the standard deviation assumed to be known

To describe the z-test we will use a simple (constructed) example.

Example 5 Has a new method of teaching made a difference?

For many years a teacher has been using the same method of teaching children to read. The scores the children obtain on a reading test have a mean of 54.6 and a standard deviation of 8.3. These values will be taken to be the population mean and the population standard deviation under the old method of teaching. The teacher tries a new method with her current class of 34 children, and their average score on the reading test is 58.1. She wants to test whether random variation underlies the difference between the average of this class (58.1) and the long-term average of previous classes (54.6), or whether there is a genuine difference.

The null and alternative hypotheses are:

H_0: The old method and new method of teaching children to read are equally effective.

H_1: The old method and new method of teaching children to read differ in their effectiveness.

If μ denotes the mean reading score of children taught by the new method, we can recast these hypotheses as

H_0: $\mu = 54.6$

H_1: $\mu \neq 54.6$.

The sample mean, $\overline{x}$, is based on the performances of $n = 34$ children. Hence, its sampling distribution is approximately normal, as a sample size of 34 is quite large. Moreover, as shown in Section 4:

- the mean of the sampling distribution of $\overline{x}$ is equal to μ
- the standard deviation of the sampling distribution of $\overline{x}$ (i.e. the standard error of $\overline{x}$) is equal to $\sigma/\sqrt{n}$.

Now, to perform a hypothesis test based on $\overline{x}$, the sampling distribution under which we calculate probabilities is the sampling distribution of $\overline{x}$ *assuming that the null hypothesis, H_0, is true*.

In the present case, if H_0 is true, then $\mu = 54.6$ and the distribution of $\overline{x}$ is approximately normal with mean 54.6 and standard deviation $\sigma/\sqrt{n}$. We know that n equals 34 but need to know the value of σ. For this example, we shall assume that the population standard deviation of scores with the new method is the same as with the old method, so $\sigma = 8.3$. All told,

$$A = 54.6, \quad \overline{x} = 58.1, \quad n = 34, \quad \sigma = 8.3.$$

Now, from the end of Section 4, if the sampling distribution of $\overline{x}$ is approximately normal with mean μ and standard deviation SE $= \sigma/\sqrt{n}$, then the distribution of the variable

$$z = \frac{\overline{x} - \mu}{\text{SE}}$$

is (approximately) the standard normal distribution (with mean zero and standard deviation one). Thus, if H_0 is true, so that $\mu = A = 54.6$, the distribution of the variable

$$z = \frac{\overline{x} - 54.6}{\text{SE}}$$

is (approximately) the standard normal distribution.

The variable z is the test statistic for the z-test. Its numerical value in this example is

$$z = \frac{\overline{x} - A}{\text{SE}} = \frac{58.1 - 54.6}{8.3/\sqrt{34}} \simeq 2.46.$$

The main result that we have obtained so far is summarised in the following box.

Test statistic and its sampling distribution when H_0 is true and σ is assumed known

For a one-sample z-test, when H_0: $\mu = A$ is true, the test statistic,

$$z = \frac{\overline{x} - A}{\text{SE}},$$

follows (approximately) the standard normal distribution, where

SE $= \sigma/\sqrt{n}$.

Activity 22 Value of z

Calculate the value of the test statistic z for the test of

H_0: $\mu = 120$

H_1: $\mu \neq 120$,

when $n = 100$, $\overline{x} = 112$, and $\sigma = 15$.

Critical values and critical regions

If the null hypothesis, H_0, is true, then z should follow the standard normal distribution. This distribution has a mean of 0, so if the value of z given by our data was very large in size (positive or negative), it would suggest that H_0 is false. The idea, then, is to reject H_0 if the observed value of z is 'too extreme' and therefore unlikely. Notice that 'too extreme' covers both large positive values and large negative values, in line with H_1, which specifies $\mu \neq A$, 'in either direction' away from A. If we cannot believe that the observed z is an observation from a standard normal distribution, then we cannot believe H_0.

We have calculated the value of the test statistic z that is given by our data. Suppose now that the test is to be performed at the 5% significance level. As discussed in Subsection 4.1 of Unit 6, H_0 will be rejected at the 5% significance level if z is in the most extreme 5% of values under the sampling distribution that applies if H_0 is true. This 'most extreme' region is the **critical region** of the test. (In this case it is the critical region at the 5% significance level.) Because of the discussion in the previous paragraph, the critical region consists of two parts: one part comprises the most extremely high 2.5% of values under the standard normal distribution, and the other part comprises the most extremely low 2.5% of values under the standard normal distribution.

The values defining the 'inner ends' of the critical region are the **critical values**. The critical values for the z-test at the 5% significance level are 1.96 and -1.96. Figure 32 shows the critical values and critical region pictorially.

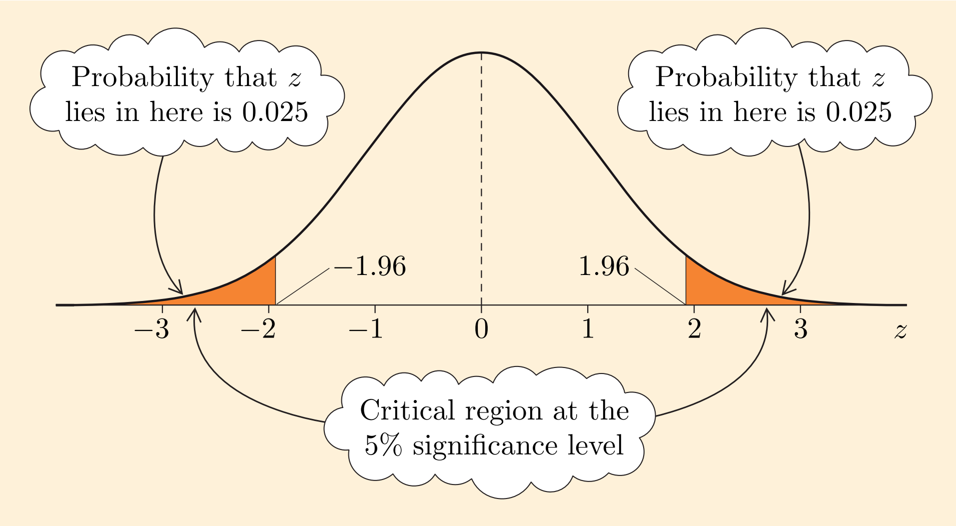


Figure 32 The standard normal distribution with the critical region and critical values (1.96 and -1.96) shown for a test at the 5% significance level

Instead of using the 5% significance level for the hypothesis test, we might want to perform the test at the more stringent 1% significance level. To do this, all that changes is the values of the critical values and hence the critical region. The critical values become 2.58 and -2.58, and the critical region is rather smaller: see Figure 33.

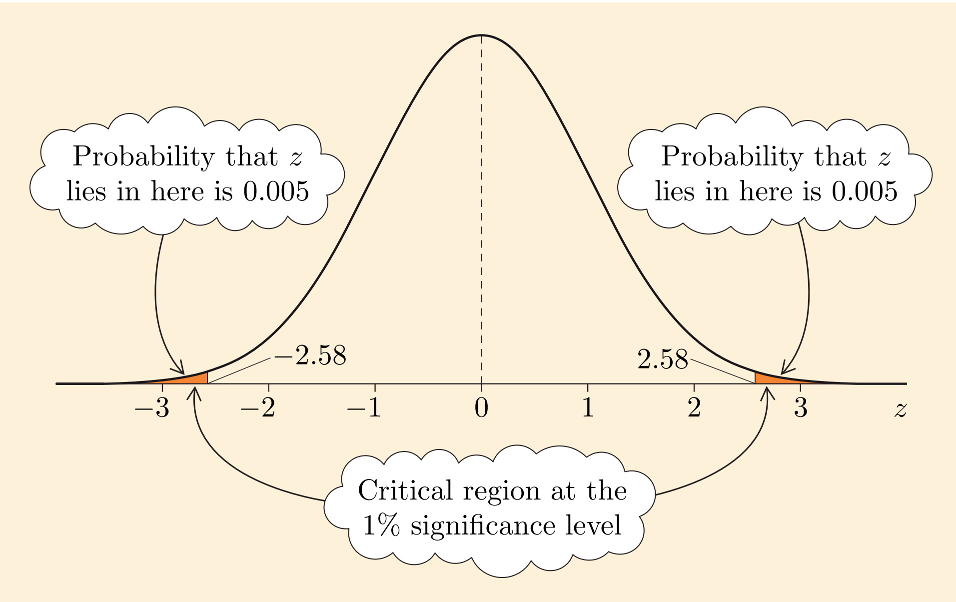


Figure 33 The standard normal distribution with the critical region and critical values (2.58 and -2.58) shown for a test at the 1% significance level

The procedure to be followed to complete the z-test is as follows.

Completing the z-test

If $z \geq 2.58$ or $z \leq -2.58$, reject H_0 at the 1% significance level.

If $1.96 \leq z < 2.58$ or $-2.58 < z \leq -1.96$, reject H_0 at the 5% significance level but not at the 1% significance level.

If $-1.96 < z < 1.96$, do not reject H_0 at the 5% significance level.

Activity 23 Where is z in marginal circumstances?

On a sketch of the standard normal distribution, show where the value of z must lie in the marginal case where H_0 is rejected at the 5% significance level but not at the 1% significance level.

As noted in Subsection 5.1 of Unit 6, we conclude that there is strong evidence against the null hypothesis if we reject H_0 at the 1% significance level. If we reject it at the 5% significance level but not the 1% level, we conclude that there is moderate (but not strong) evidence against the null hypothesis. If we do not reject H_0 at the 5% significance level, we have, in the words of Subsection 5.1 of Unit 6, either 'little' or 'weak' evidence against H_0.

We will use 'strong' whenever we reject H_0 at the 1% significance level in this unit; the evidence might in fact be 'very strong', but we will not be testing H_0 at the 0.1% level.

Example 6 Completing the z-test started in Example 5

In Example 5, the test statistic (the data z-value) takes the value 2.46. This value exceeds 1.96 but not 2.58. We conclude that there is moderate evidence that the old and new methods are not equally effective at teaching children to read.

As the new method gave an average score of 58.1, while the average under the old method was 54.6, and a higher test score means better reading ability, there is moderate evidence that the new method is better than the old method.

Activity 24 A z-test in manufacturing

A firm is engaged in putting finishes on work surfaces for kitchen manufacturers. Previously, the work was done in very large batches, so the time spent setting up the machine did not affect production too much. However, with a change in the pattern of demand the batch size has had to be considerably reduced, so the time spent setting the machine to different specifications is becoming more important.

Last year the manufacturing manager found that the machine setting had been changed very many times and the mean time taken for a change was 26.1 minutes. The operators suggested a way in which the set-up time might be reduced, but the manager was unconvinced and feared that the set-up time might actually be increased. Nevertheless, it was agreed to try out this new method for two weeks. A z-test would then be performed to examine whether or not the mean time for setting up under the new method differs from the mean time taken last year.

In the two-week testing period, the machine was reset on 53 occasions, taking a mean time of 20.9 minutes.

(a) What are the appropriate null and alternative hypotheses?

(b) Give the values of A, $\overline{x}$ and n.

(c) Assume that the standard deviation, σ, equals 12.3. Calculate the value of the test statistic.

(d) Is the null hypothesis rejected at the 5% significance level? Is it rejected at the 1% significance level?

(e) What do you conclude from the hypothesis test?

5.2 The z-test with unknown standard deviation

In Subsection 5.1, we developed one-sample z-tests under the assumption that σ is known. Now σ is the standard deviation of the population from which the sample data are drawn. Typically its actual value will not be known, but if we have a large sample then the sample standard deviation, s, provides a good estimate of σ. Moreover, provided the sample size is large, the one-sample z-test can be performed with σ replaced by s. Specifically, we calculate the *estimated standard error* (ESE) of $\overline{x}$,

Compare formula for ESE with $\text{SE} = \sigma/\sqrt{n}$.

$$\text{ESE} = \frac{s}{\sqrt{n}},$$

and put

$$z = \frac{\overline{x} - A}{\text{ESE}}.$$

You might be slightly disquieted by the bald assertion that, for large samples, replacing SE by its estimated value ESE makes no difference to the (approximate) standard normal distribution of $(\overline{x} - A)/\text{SE}$. After all, ESE is not the correct quantity to divide by; SE is. It is the assumption of a large sample that saves the day. In Unit 10 we give tests for small samples (t-tests) which take the difference between ESE and SE into account. Differences between those tests and z-tests are small when the sample size is above about 25.

At the end of Section 2, it was asserted that the sampling distribution of the mean will always be approximately normal for sample sizes greater than 25. It was also argued that the sampling distribution of the mean will actually be approximately normal for sample sizes (much) smaller than $n = 25$ for many population distributions. In that sense, the notion of $n = 25$ being large enough errs on the 'careful' side. When SE is replaced by its estimated value (ESE), however, a sample size of 25 is only just enough for a z-test to be usable. We will continue to use this 'rule of thumb', but $n = 25$ is no longer a 'generous' value – many would prefer to use z-tests only for samples that are a bit larger than that.

What is a large enough sample for a z-test?

As a rough guide you can assume that, whatever the population distribution, for sample sizes greater than 25, the z-test is applicable.

As this jolly logo shows, ESE also stands for ‘Exceptional Student Education’ . . . an educational program in schools in Citrus County, Florida, USA

The next two boxes lay out the full requirements and procedure for the one-sample z-test. They cover both the cases where σ is known and where it must be estimated. The first box gives the key pieces of information that you should pick out for a z-test when you are reading details about a survey or experiment.

Key values for a one-sample z-test

The information you need to know for a one-sample z-test is:

- the hypothesised population mean (A) under the null hypothesis
- the sample mean $(\overline{x})$
- the sample size (n)
- the population standard deviation (σ), or a good estimate of σ.

Procedure: the one-sample z-test

1. Set up the null and alternative hypotheses,

   H_0: $\mu = A$

   H_1: $\mu \neq A$,

   where μ is the population mean.
2. Calculate the test statistic, z:
   - If the population standard deviation (σ) is known,

     $$z = \frac{\overline{x} - A}{\text{SE}}, \quad \text{where SE} = \frac{\sigma}{\sqrt{n}}.$$
   - If σ is unknown but the sample size (n) is 25 or more,

     $$z = \frac{\overline{x} - A}{\text{ESE}}, \quad \text{where ESE} = \frac{s}{\sqrt{n}}.$$

   Here $\overline{x}$ is the sample mean and s is the standard deviation of the sample. SE is the standard error of the mean and ESE is the estimated standard error.
3. Compare z with the appropriate critical values, which are 1.96 and -1.96 at the 5% significance level and 2.58 and -2.58 at the 1% significance level.

- If $z \geq 2.58$ or $z \leq -2.58$, then H_0 is rejected at the 1% significance level.
- If $1.96 \leq z < 2.58$ or $-2.58 < z \leq -1.96$, then H_0 is rejected at the 5% significance level but not at the 1% significance level.
- If $-1.96 < z < 1.96$, then H_0 is not rejected at the 5% significance level.

4. State the conclusions that can be drawn from the test.

We are now in a position to start answering some of the questions we asked about the BCS survey data in Subsection 1.3. The investigation illustrates use of the one-sample z-test when σ is unknown.

Example 7 Reading scores of 7-year-old children in BCS survey

In a question that was posed at the end of Subsection 1.3, we asked whether the sample of children from the BCS survey in 2004–2005 could be considered to have come from the population of children for whom the British Ability Scales reading score was developed. The overall population mean reading scores for British children are taken to be 96 for 7-year-old children. We wrote down the following null and alternative hypotheses:

H_0: For British children aged 7 in 2004–2005, the mean reading score is equal to 96

H_1: For British children aged 7 in 2004–2005, the mean reading score is not equal to 96.

We can recast these hypotheses as

H_0: $\mu = 96$

H_1: $\mu \neq 96$,

where μ is the population mean of the reading scores of all British 7-year-old children in 2004–2005. The data from the BCS concerning 7-year-old children are summarised in Table 3.

Table 3 Further summary statistics for data on reading scores of 7-year-old children

Sample size	Sample mean	Sample standard deviation
396	111.28	26.668

(This data is copyright and owned by the Economic and Social Data Service.)

Although σ is unknown, the sample size is considerably greater than 25, so the ESE may be used in calculating z. Thus the information required for the z-test is:

$$A = 96, \quad \overline{x} = 111.28, \quad n = 396, \quad s = 26.668.$$

We can now calculate the test statistic:

$$z = \frac{\overline{x} - A}{\text{ESE}} = \frac{\overline{x} - A}{s/\sqrt{n}} = \frac{111.28 - 96}{26.668/\sqrt{396}} \simeq 11.40.$$

Assuming that the null hypothesis, H_0, is true, 11.40 is a value from the standard normal distribution. However, 11.40 is much bigger than the 1% critical value of 2.58. Hence the z-test clearly rejects H_0 at the 1% level.

We conclude that there is strong evidence that the mean reading score for 7-year-old children in 2004–2005 is not equal to the overall mean reading score

for 7-year-old children. At face value, this is a little surprising – there seems no obvious factor to cause a difference in reading ability between the 7-year-old children in the 2004–2005 BCS survey and the 7-year-olds in the population of British children for whom the reading test was originally developed. Given that the mean reading score for the 7-year-old children in the BCS survey is larger than the overall mean reading score, for some reason the BCS children seem to have performed rather better than expected (on average).

The following two activities provide you with practice in applying the z-test. The first one continues our investigation of the BCS data. It concerns the reading scores of 8-year-old children. The second concerns some data on earnings.

Activity 25 Reading scores of 8-year-old children in BCS survey

In the BCS investigation, the following results were obtained for 8-year-old children.

Table 4 Summary statistics for data on reading scores of 8-year-old children

Sample size	Sample mean	Sample standard deviation
283	126.92	27.711

(This data is copyright and owned by the Economic and Social Data Service.)

The overall mean reading score for 8-year-old children is 116.

Carry out a z-test to investigate whether the sample of 8-year-old children was selected from a population whose mean reading score is equal to the overall mean score for 8-year-old children. Comment on your result.

Activity 26 Wages of female employees

A random sample of 810 female local government clerical officers and assistants had a mean wage of \$373.40 per week in 2011 with a standard deviation of \$138.20. The overall mean weekly wage for female employees in 2011 was \$381.50. (Source: *Annual Survey of Hours and Earnings*, 2011.) Investigate whether the mean weekly wage of female local government clerical officers and assistants differed from the overall mean weekly wage for female employees in 2011. Comment on your result.

You have now covered the material related to Screencast 5 for Unit 7 (see the M140 website).

Exercises on Section 5

Exercise 16 Reading scores of 7-year-old girls in BCS survey

In the BCS investigation, the following results were obtained for 7-year-old girls. (These results have been extracted from Table 2 in Subsection 1.3.)

Table 5 Summary statistics for data on reading scores of 7-year-old girls

Sample size	Sample mean	Sample standard deviation
190	113.42	25.464

The overall mean reading score for 7-year-old children is 96.

Carry out a z-test to investigate whether the sample of 7-year-old girls was selected from a population whose mean reading score is equal to the overall mean score for 7-year-old children. Comment on your result.

Exercise 17 Weight of pigs

A random sample of 533 pigs of a certain breed that had been fed a special diet were weighed. They had a mean weight of 81.92 kg with a standard deviation of 15.65 kg. The mean weight of this breed of pig when fed the standard diet is 80 kg. Evaluate the evidence that the special diet changes the mean weight of this breed of pig.

Exercise 18 An exciting exercise: paint drying

A consumer magazine, when comparing various brands of paint, stated that the drying time of one particular brand was exactly four hours. The manufacturers of that paint were not particularly pleased with this as they believed the drying time for their paint was shorter. They organised a trial in which the paint was tested by a random sample of 40 customers, all of whom were decorating their living rooms. For this sample the mean drying time was found to be 3.80 hours and the standard deviation was 0.55 hours.

(a) Analyse the sample data to test whether the drying time given by the consumer magazine is correct.

(b) What reservations might there be about your conclusion?

In 2011/12, the internet (including one national newspaper) was abuzz with news of the forthcoming inaugural World Watching Paint Dry Championships to held in Stoke-on-Trent in July 2012. Competitors were to each be given a one-metre square patch of freshly emulsioned wall at which to stare as it slowly dried. There were said to be 42 entrants, from the UK, USA, India and Hungary. Unfortunately, there is no evidence that the event actually took place.

6 The two-sample z-test

In this section we develop the **two-sample z-test**, which is used to analyse the difference in locations between two populations. There were plenty of examples of this raised in the context of the BCS and its data on reading scores in Subsections 1.2 and 1.3. One such question posed there was:

> *For British children aged 7 in 2004–2005, did boys' and girls' reading scores differ in location?*

Here, the two populations which we wish to compare in terms of their reading abilities are the population of British boys aged 7 in 2004–2005 and the population of British girls aged 7 in 2004–2005. Another example is the question

> *For British children aged 7–8 in 2004–2005, did reading scores differ in location according to their father's occupation?*

Here, the two populations which we wish to compare in terms of their reading abilities are the population of British children aged 7–8 in 2004–2005 whose father's occupation was coded 1 in Table 1 (managerial, technical, professional and skilled non-manual occupations) and the population of British children aged 7–8 in 2004–2005 whose father's occupation was coded 2 (skilled manual, partly skilled and unskilled occupations).

As in Section 5, comparisons will be made using hypothesis tests comparing means, and the two-sample z-test will be appropriate when both samples are large. To develop this test we use the reading scores from the BCS sample. We examine the first of the above questions:

> *For British children aged 7 in 2004–2005, did boys' and girls' reading scores differ in location?*

The following are appropriate null and alternative hypotheses:

H_0: For British children aged 7 in 2004–2005, the mean reading score for girls is equal to the mean reading score for boys

H_1: For British children aged 7 in 2004–2005, the mean reading score for girls is not equal to the mean reading score for boys.

We shall now introduce some symbols that will enable us to express our hypotheses more concisely and will also be helpful in explaining a theoretical result that we need. We are investigating two populations of values: the reading scores of all British 7-year-old girls in 2004–2005 and the reading scores of all

The subscripts 'g' and 'b' always relate to girls and boys, respectively.

British 7-year-old boys in 2004–2005. We shall let the means of these two populations be μ_g and μ_b, and the standard deviations be σ_g and σ_b. It is worth noting that the values of these quantities cannot be known: not all British 7-year-old girls and boys actually took this test in 2004–2005. So there is no way we could actually calculate μ_g, μ_b, σ_g and σ_b, but they enable us to make precise statements.

For a start, we can use μ_g and μ_b to write the hypotheses concisely as

$$H_0\colon \mu_\mathrm{g} = \mu_\mathrm{b}$$
$$H_1\colon \mu_\mathrm{g} \neq \mu_\mathrm{b},$$

or, equivalently, as

$$H_0\colon \mu_\mathrm{g} - \mu_\mathrm{b} = 0$$
$$H_1\colon \mu_\mathrm{g} - \mu_\mathrm{b} \neq 0.$$

This last form is the one we shall actually use to derive the test statistic.

Although we do not know test values for all children, the values for the samples of girls and boys in the BCS are known. We shall denote these samples' sizes by n_g and n_b, the sample means by $\overline{x}_\mathrm{g}$ and $\overline{x}_\mathrm{b}$, and the sample standard deviations by s_g and s_b. Their values were set out in Table 2 (Subsection 1.3), but we do not need them at the moment.

As we have expressed our null hypothesis as $\mu_\mathrm{g} - \mu_\mathrm{b} = 0$, it seems intuitively sensible to test the hypothesis by looking at the difference between the sample means, $\overline{x}_\mathrm{g} - \overline{x}_\mathrm{b}$. Before we can develop our hypothesis test, we need a theoretical result about the sampling distribution of the difference between two sample means.

You already know, from Section 4, that, because n_g and n_b are large, the sampling distribution of $\overline{x}_\mathrm{g}$ is approximately normal with mean μ_g and standard error $\sigma_\mathrm{g}/\sqrt{n_\mathrm{g}}$, and similarly that the sampling distribution of $\overline{x}_\mathrm{b}$ is approximately normal with mean μ_b and standard error $\sigma_\mathrm{b}/\sqrt{n_\mathrm{b}}$. We may conceive the first of these sampling distributions by thinking of all the possible samples of size n_g that we could select from the population of scores of all 7-year-old girls. We then imagine that we could calculate $\overline{x}_\mathrm{g}$ for each of these samples and look at their distribution. Similar considerations apply to the sampling distribution of $\overline{x}_\mathrm{b}$.

Now, think of all the possible means $\overline{x}_\mathrm{g}$ of samples of size n_g of girls and also all the possible means $\overline{x}_\mathrm{b}$ of samples of size n_b of boys. If we select just one value of $\overline{x}_\mathrm{g}$ and one value of $\overline{x}_\mathrm{b}$, we can calculate the difference $\overline{x}_\mathrm{g} - \overline{x}_\mathrm{b}$. Now think of all the possible pairs of values $\overline{x}_\mathrm{g}$ and $\overline{x}_\mathrm{b}$ we could select, and suppose we calculate $\overline{x}_\mathrm{g} - \overline{x}_\mathrm{b}$ for each of them. Then the distribution of all these differences is the **sampling distribution of the difference between two means**.

We require three results that are known about this sampling distribution. First, the mean of the sampling distribution of $\overline{x}_\mathrm{g} - \overline{x}_\mathrm{b}$ is equal to $\mu_\mathrm{g} - \mu_\mathrm{b}$, as you might expect. The second result requires the two samples to be independent of each other – here that is clearly the case, as the choice of girls was completely separate from the choice of boys. As long as the samples are independent, the standard deviation of the sampling distribution is given by

$$\mathrm{SE} = \sqrt{\frac{\sigma_\mathrm{g}^2}{n_\mathrm{g}} + \frac{\sigma_\mathrm{b}^2}{n_\mathrm{b}}},$$

and this standard deviation is called the **standard error of the difference between two means**. Notice that it is larger than the standard errors of $\overline{x}_\mathrm{g}$ and $\overline{x}_\mathrm{b}$, which are $\sigma_\mathrm{g}/\sqrt{n_\mathrm{g}}$ and $\sigma_\mathrm{b}/\sqrt{n_\mathrm{b}}$, respectively. This is because we are looking at the difference between two sample means; both means can vary, so there is

more variation in the difference between them. Notice also that the standard error of the difference between two means is neither the sum nor the difference of the standard errors of the individual means. The next box summarises these results.

Mean and standard deviation of the sampling distribution of the difference between two means

- The mean of the sampling distribution is equal to $\mu_{\text{g}} - \mu_{\text{b}}$, the difference between the population means.
- The standard deviation of the sampling distribution is called the standard error of the difference between two means, and is given by

  $$\text{SE} = \sqrt{\frac{\sigma_{\text{g}}^2}{n_{\text{g}}} + \frac{\sigma_{\text{b}}^2}{n_{\text{b}}}},$$

  where n_{g} and n_{b} are the sizes of the samples, and σ_{g} and σ_{b} are the population standard deviations.

Furthermore, provided the sample sizes are sufficiently large, the sampling distribution of the differences between two sample means is approximately normal. This is the third result that we require.

Approximate normality of the sampling distribution of the difference between two means

If n_{g} and n_{b} are large, no matter what shape the population distributions, the sampling distribution of the difference between two means based on samples of sizes n_{g} and n_{b} will in practice be approximately normal.

From these results, $\overline{x}_{\text{g}} - \overline{x}_{\text{b}}$ is approximately normally distributed with mean $\mu = \mu_{\text{g}} - \mu_{\text{b}}$ and standard deviation $\sigma = \text{SE}$. Thus, the formula given in Subsection 3.3 can be used to transform $\overline{x}_{\text{g}} - \overline{x}_{\text{b}}$ to a quantity which follows (approximately) the standard normal distribution:

$$z = \frac{x - \mu}{\sigma} = \frac{(\overline{x}_{\text{g}} - \overline{x}_{\text{b}}) - (\mu_{\text{g}} - \mu_{\text{b}})}{\text{SE}}.$$

Now to obtain our test statistic, we assume that the null hypothesis H_0 is true, so $\mu_{\text{g}} - \mu_{\text{b}} = 0$. We still cannot calculate z, as we do not know σ_{g} and σ_{b}. We deal with this problem exactly as we did in Subsection 5.1, by replacing σ_{g} by s_{g} and σ_{b} by s_{b}. This leads to the estimated standard error of $\overline{x}_{\text{g}} - \overline{x}_{\text{b}}$:

$$\text{ESE} = \sqrt{\frac{s_{\text{g}}^2}{n_{\text{g}}} + \frac{s_{\text{b}}^2}{n_{\text{b}}}}.$$

This logo suggests a more relaxing form of ESE

Test statistic and its sampling distribution when H_0 is true

For a two-sample z-test, when H_0: $\mu_{\text{g}} - \mu_{\text{b}} = 0$ is true, the test statistic,

$$z = \frac{\overline{x}_{\text{g}} - \overline{x}_{\text{b}}}{\text{ESE}}, \quad \text{where ESE} = \sqrt{\frac{s_{\text{g}}^2}{n_{\text{g}}} + \frac{s_{\text{b}}^2}{n_{\text{b}}}},$$

follows (approximately) the standard normal distribution.

For the one-sample z-test, we used the rule of thumb that the sample size had to be at least 25. To justify use of a two-sample z-test, we apply this rule of thumb to both samples and require that each sample size should be at least 25.

Since the test statistic above has the standard normal distribution (approximately) when the null hypothesis is true, the critical values are exactly the same as those in Subsection 5.1 for a one-sample hypothesis test. We can reject H_0 at the 1% significance level if $z \geq 2.58$ or if $z \leq -2.58$, and we can reject H_0 at the 5% significance level if $z \geq 1.96$ or $z \leq -1.96$. Otherwise we cannot reject H_0.

Example 8 Comparing the mean reading scores of girls and boys

We are now able to perform the two-sample z-test with which the current subsection was introduced. The hypotheses are:

$$H_0: \mu_{\text{g}} = \mu_{\text{b}}$$
$$H_1: \mu_{\text{g}} \neq \mu_{\text{b}},$$

where μ_{g} is the population mean reading score for 7-year-old girls in 2004–2005, and μ_{b} is the population mean reading score for 7-year-old boys in 2004–2005. The data on which the test will be based were given as Table 2 (Subsection 1.3) and are repeated in Table 6.

Table 6 Summary statistics for data on reading scores of 7-year-old children

	Sample size	Sample mean	Sample standard deviation
Boys	206	109.31	27.671
Girls	190	113.42	25.464

(This data is copyright and owned by the Economic and Social Data Service.)

Key values for a two-sample z-test

In general, call the two groups A and B. The information you need to know for a two-sample z-test is:

- the sample means ($\overline{x}_A$ and $\overline{x}_B$)
- the sample sizes (n_A and n_B)
- the population standard deviations (σ_A and σ_B), or good estimates of them (s_A and s_B).

In this example we are using ‘g’ and ‘b’ to distinguish the two groups, rather than A and B. We have:

$$\overline{x}_{\text{g}} = 113.42, \quad \overline{x}_{\text{b}} = 109.31, \quad n_{\text{g}} = 190, \quad n_{\text{b}} = 206,$$
$$s_{\text{g}} = 25.464, \quad s_{\text{b}} = 27.671.$$

Both $n_{\text{g}} = 190$ and $n_{\text{b}} = 206$ are greater than 25, so we can assume that the z-test is applicable.

In the two-sample case, it is easier to calculate the value of z in two stages.

We first calculate the value of ESE, the estimated standard error of $\overline{x}_{\text{g}} - \overline{x}_{\text{b}}$:

$$\text{ESE} = \sqrt{\frac{s_{\text{g}}^2}{n_{\text{g}}} + \frac{s_{\text{b}}^2}{n_{\text{b}}}} = \sqrt{\frac{25.464^2}{190} + \frac{27.671^2}{206}} \simeq 2.670.$$

Hence the value of the test statistic is

$$z = \frac{\overline{x}_{\text{g}} - \overline{x}_{\text{b}}}{\text{ESE}} \simeq \frac{113.42 - 109.31}{2.670} \simeq 1.54.$$

The critical values are 1.96, -1.96 (5%) and 2.58, -2.58 (1%). Since $-1.96 < 1.54 < 1.96$, we cannot reject H_0 at the 5% significance level. There is little evidence to suggest that the mean reading scores in 2004–2005 for 7-year-old boys and girls were different.

The procedure for the two-sample z-test is summarised in the following box.

Procedure: two-sample z-test

1. Set up the null and alternative hypotheses,

   H_0: $\mu_A = \mu_B$

   H_1: $\mu_A \neq \mu_B$,

   where μ_A and μ_B are the means of populations A and B, respectively.
2. Calculate the test statistic

   $$z = \frac{\overline{x}_A - \overline{x}_B}{\text{ESE}},$$

   where the estimated standard error of $\overline{x}_A - \overline{x}_B$ is

   $$\text{ESE} = \sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}.$$

   Here, n_A and n_B are the sample sizes of random samples from populations A and B respectively, $\overline{x}_A$ and $\overline{x}_B$ are the sample means, and s_A and s_B are the sample standard deviations.
3. Compare z with the appropriate critical values, which are 1.96 and -1.96 at the 5% significance level, and 2.58 and -2.58 at the 1% significance level.
   - If $z \geq 2.58$ or $z \leq -2.58$, then H_0 is rejected at the 1% significance level.
   - If $1.96 \leq z < 2.58$ or $-2.58 < z \leq -1.96$, then H_0 is rejected at the 5% significance level but not at the 1% significance level.
   - If $-1.96 < z < 1.96$, then H_0 is not rejected at the 5% significance level.
4. State the conclusions that can be drawn from the test.

In the two-sample z-test, it doesn't actually matter which of the two groups of interest you label A and which B. If you swapped the roles of A and B over, you would change the sign of z but nothing else. In particular, the conclusions of the test would be the same in either case.

Activity 27 Mean reading scores of girls and boys at age 8

In the BCS investigation, the following results were obtained for 8-year-old children.

Table 7 Summary statistics for data on reading scores of 8-year-old children

	Sample size	Sample mean	Sample standard deviation
Boys	145	126.38	29.927
Girls	138	127.49	25.064

(This data is copyright and owned by the Economic and Social Data Service.)

Carry out a two-sample z-test to investigate whether the mean reading score of 8-year-old girls in 2004–2005 was equal to the mean reading score of 8-year-old boys in 2004–2005. Comment on your result.

You might have noticed something interesting about the results for 7-year-old and 8-year-old children. For the younger children, the girls' sample mean score was $113.42 - 109.31 = 4.11$ more than that for boys, whereas for the older children the girls' sample mean score was $127.49 - 126.38 = 1.11$ higher. One might have thought at first glance that there was an interesting effect here: at the younger age, girls are ahead of boys in reading ability, but a year later boys seem to be catching up. Not so, however: our hypothesis tests showed that in neither case was there any evidence of a real difference, or therefore, of any such effect. The differences in the samples that we observed can easily have arisen by chance.

Does the level of education of parents have an affect on the reading scores of their children? In the next activity you will investigate this in the context of the BCS survey. This study classified parental education into two categories: those who finished full-time education by age 16 and those who continued after 16 (see Table 1, Subsection 1.2).

Activity 28 Mean reading scores according to parental education

This activity addresses another of the questions raised in Subsection 1.2:

> *For British children aged 7–8 in 2004–2005, did reading scores differ in location according to the level of their parents' education?*

Table 8 provides the relevant summary data from the BCS.

Table 8 Summary statistics for data on reading scores and parental education

Parental education	Sample size	Sample mean	Sample standard deviation
Ended by age 16	389	116.12	28.775
Continued after age 16	199	123.15	24.603

(Note that 679 children were tested for their reading ability, but no information was available on when the parents of 91 of the children completed their education.)

Carry out a hypothesis test to investigate whether children whose parents' education continued beyond age 16 scored differently on average on the reading test from those children whose parents' education ended by age 16.

You might have expected the answer to Activity 28 before analysing the data. That is, denoting μ_{C} as the mean reading score of children whose parental education continued after age 16 and μ_{E} as the mean reading score of children whose parental education ended by age 16, you might have thought of doing the following: testing the null hypothesis that $\mu_{\mathrm{C}} = \mu_{\mathrm{E}}$ with the purpose of seeing whether, as you suspect, μ_{C} is actually greater than μ_{E}, disregarding the possibility that μ_{C} could be less than μ_{E}. Hypothesis tests undertaken when a particular type of inequality between the two groups is of interest are the

one-sided tests mentioned in a margin note at the start of Section 5 and to be looked at briefly in Unit 10.

You have now covered the material related to Screencast 6 for Unit 7 (see the M140 website).

Exercises on Section 6

Exercise 19 Mean reading scores according to fathers' occupations

This exercise concerns another question posed in Subsection 1.2, namely:

> *For British children aged 7–8 in 2004–2005, did reading scores differ in location according to their fathers' occupations?*

Table 9 provides the relevant summary data from the BCS. Note that, as in Table 1 (Subsection 1.2), '1' denotes 'managerial, technical, professional and skilled non-manual' occupations while '2' denotes 'skilled manual, partly skilled and unskilled' occupations.

Table 9 Summary statistics for data on reading scores and father's occupation

Father's occupation	Sample size	Sample mean	Sample standard deviation
1	316	120.55	24.221
2	203	117.17	30.085

(No information was available on father's occupation for 160 individuals.)

Carry out a two-sample hypothesis test to investigate whether the mean score of children whose father had an occupation coded 1 differs from that of children whose father had an occupation coded 2. Comment on your result.

Exercise 20 Calcium for babies

This exercise is related to an investigation of the effect of vitamin D supplementation for the prevention of low levels of calcium in newborn babies. The data given in Table 10 come from a clinical trial in which a sample of babies who were breast-fed were compared with a sample of babies who were bottle-fed: the measured quantity was the level of calcium in the baby's blood ('serum calcium') at 1 week of age.

Table 10 Summary statistics for data on serum calcium for week-old babies

	Sample size	Sample mean	Sample standard deviation
Breast-fed	64	2.45	0.292
Bottle-fed	169	2.30	0.274

(Source: Cockburn et al. (1980) 'Maternal vitamin D intake and mineral metabolism in mothers and their newborn infants', *British Medical Journal*, vol. 281, pp. 11–14)

Carry out a two-sample z-test to investigate whether the mean serum calcium level of week-old babies was the same whether they were breast-fed or bottle-fed.

Exercise 21 Peak flow rate of lungs

The peak flow rate is a measure of how well a person's lungs are functioning. It is the maximum rate in litres per minute at which air can be expelled through a peak flow meter. In an investigation of the possibility that chronic bronchitis, although a disease of adult life, starts in childhood, the peak flow rates of a large number of school children without persistent coughs were measured. Amongst other details recorded were whether the child lived in an urban or a rural area. Data for urban and rural areas are summarised in Table 11. Use a two-sample z-test to examine whether the average peak flow rate of children differs in these two groups.

Table 11 Peak flow rates for children without persistent coughs

	Sample size	Sample mean	Sample standard deviation
Urban	485	226	52
Rural	637	231	53

(Source: unpublished data collected by Professor J.R.T. Colley, University of Bristol)

7 Computer work: one-sample z-tests

In this section you will use Minitab to perform one-sample z-tests. These are similar to the tests you have performed earlier in this unit, except that Minitab gives the results of hypothesis tests in terms of p-values, while in earlier sections we have only considered specific significance levels (5% and 1% significance levels). The use of p-values with sign tests was explained in Unit 6. Their use with z-tests is identical, but is described explicitly in the Computer Book.

You should now turn to the Computer Book and work through Chapter 7. The chapter starts with the interactive computer resources connected with Section 3 of this unit; you should do them now if you have not already done so. You should then do the Minitab work that is contained in the rest of Chapter 7.

8 Conclusions and reservations

We have answered many of the questions raised in Section 1, and we have learned a lot about children's reading ability and factors affecting it, at any rate for British children aged 7 and 8 in 2004–2005. We summarise our conclusions below. As usual, though, after coming to such conclusions, we should stop and look for reservations that might arise.

- Are there any problems with the data that might throw doubt on conclusions drawn from them?
- Were appropriate statistical methods used in analysing the data?

We shall look at both these questions. To address the second question we shall discuss when z-tests should be used. We then note limitations on the way conclusions are stated and interpreted.

Conclusions

We began this unit by asking the general question:

> *What factors affect a child's reading ability*?

In Section 1, we refined this question to produce several more specific questions that we could attempt to answer using BCS data. In Sections 5 and 6, we carried out hypothesis tests that related to these questions. All these tests involved hypotheses about the population from which the BCS sample was drawn, that of British children aged 7 and 8 in 2004–2005.

In Example 7 (Subsection 5.2), we found that we could reject the null hypothesis that 7-year-old British children in 2004–2005 had the overall population mean reading score for 7-year-olds. Similarly, in Activity 25 (Subsection 5.2), we found that we could reject the null hypothesis that 8-year-old British children in 2004–2005 had the overall population mean reading score for 8-year-olds. (A related result for 7-year-old girls was obtained in Exercise 16 in Section 5.)

In Example 8 (Section 6), we found that, for 7-year-old children, the null hypothesis that the population mean for boys was equal to that for girls could not be rejected. In Activity 27 (Section 6), we also found that the same was true for 8-year-old children.

In Activity 28 (Section 6), we found strong evidence that the mean reading score was higher for children whose parents' education had lasted longer. (Something less expected happened with respect to father's occupation in Exercise 19.)

Reservations about the data

Probably the main reservation about the data is whether they can be considered a random sample from the relevant population. As was discussed in Activity 3 (Subsection 1.2), the data do not come from a formal random sample of British children aged 7 and 8 in 2004–2005, of the sort one might draw using a sampling frame and random numbers. But it might still be the case that the data can be *treated* as if they had been drawn in that way. How would a 'real' random sample differ from the BCS 2004–2005 sample? The main difference was raised in Activity 3; all the children in our sample have at least one parent who is in the BCS survey, and therefore was born in a particular week in 1970. In a true random sample of British 7- and 8-year-old children, not every child would have a parent aged 34.

There are other features of the sampling process that might lead to the sample of children being unrepresentative:

- Children could be included only if the BCS 2004–2005 investigators had managed to trace their parents. People in the original BCS sample whose lifestyles involve moving around a lot may have been harder to trace, and therefore their children would be less likely to be in the sample.
- There are missing data. This data may not be missing completely 'at random' – which might be OK, provided there is not too much of it – but its very missingness might be connected to the things you are trying to measure. (This is a common problem in real-world statistics.)

  For example, parents with less education *might* be more reluctant to say so in response to a survey, in which case children with such parents might be under-represented; worse, such parents might be more likely to not respond to the education question if they know their child is not reading especially well, and they don't want to be 'blamed' for this situation.

- The data on parental education simply give the age at which *one* of the parents left full-time education, and say nothing about which parent it was. Also, nothing is said about any qualifications he or she gained, or about any part-time study.

These reservations about the randomness/representativeness of the sample are probably less important than the reservation about the parents' age, but they should not be forgotten.

Any other reservations?

8.1 When to use the z-test

The z-test can be applied in many situations, though it does have limitations. In this subsection the characteristics of the test are described so that you can recognise when it is appropriate.

The sample size must be large

It is unnecessary to know anything about the distribution of the population from which the sample is selected, because the test is based on the fact that the sampling distribution of the mean of a sample of size n is approximately normal, provided n is sufficiently large.

As a simple rule of thumb, we assume that in the one-sample case, n should be greater than 25, and in the two-sample case, n_A and n_B should both be greater than 25. If the sample size is less than 25, you should not apply the z-test. (If you believe that the population distribution is extremely skew – which has not been the case for any distribution in this unit – then it is safer only to use the z-test if the sample size is considerably greater than 25.)

In Unit 10 you will meet another hypothesis test, the t-test, which you can apply under some circumstances when the sample size is less than 25.

The sample values should consist of numerical measurements

The z-test should be applied only to data which consist of numerical measurements. Length, weight, time, scores in a test and petrol consumption are all examples of such data. The z-test cannot be applied, for example, to data which might be coded, such as perhaps hair colour or disease type – with such data the concept of a population mean or a sample mean is not really meaningful.

The samples should be unrelated

This restriction applies only to the two-sample z-test. The samples from the two populations should be unrelated and so not consist of data collected in pairs, each pair coming from the same individual.

All the hypothesis tests that we have performed in this unit were based on data that met the requirements for the one- or two-sample z-tests. Sample sizes were above 25 (substantially so for the BCS data), sample values were numerical measurements (often scores on a reading test), and the two-sample z-test was only ever applied to unrelated samples from two separate populations.

8.2 Limitations in stating conclusions

In stating conclusions from any hypothesis test, the following factors must be borne in mind.

- A sampling error may have occurred.
- The conclusions should match the population from which the sample was drawn.
- The conclusions must not make causal statements which are not supported by the way the data arose.

Let us look at each of these briefly.

Sampling errors

You should always bear in mind that a sampling error might have occurred; that is, the result of any hypothesis test might be due to sampling variation. Hypothesis tests do not provide *proofs* of the truth of either the null or alternative hypotheses. They just attempt to assess the evidence for or against the hypotheses. For example, if the null hypothesis is rejected, that means that there is evidence against the null hypothesis, but not that the null hypothesis is definitely wrong. However, with the BCS data, in most of the hypothesis tests where we rejected the null hypothesis, the test statistic came out much higher numerically than the critical values (it easily gave 'strong evidence'), so with those tests it is unlikely – but still *possible* – that sampling error has led to erroneous conclusions.

What can we say about the populations?

A major difficulty with the BCS data is that it is not clear that these data can be treated as a random sample from *any* population. But they are clearly likely to be much more representative of the population of British children than of, say, Ugandan children. The stated conclusions were explicit in referring to British children and to the year, 2004–2005, in which the data were collected, although we should perhaps have referred to the population of British children, aged 7 or 8 in 2004–2005, *who had at least one parent aged 34*, as all these characteristics are common to the children in our sample. Nevertheless, it seems reasonably

plausible that the data would still be representative of British children in some other year close to 2004–2005, say 2003 or 2007, since reading skills are unlikely to change very rapidly. But it would be a mistake to apply the conclusions directly to the population of British children in 1970, say, or 2013.

What can we say about causal statements?

Causality will be discussed at much greater length in another hypothesis testing context in Unit 8.

Can we make any conclusions about what might have *caused* any differences for which we have evidence? For the BCS data, the answer is, essentially, 'no'! Our conclusions are not worded in causal terms; for instance, we concluded (in Activity 28) that, for British children aged 7 and 8 in 2004–2005, those whose parental education was beyond the age of 16 had a higher mean reading score than did those whose parent left education earlier. Worded like that, the conclusion says nothing about how this difference arose; but there is a great temptation to suppose that the parent's level of education *caused* the difference in mean reading score. This causal conclusion goes beyond what the data tell us. Instead, there could well be one or more other factors that underly both a child's reading ability and whether a parent of the child was educated past the age of 16. We just cannot tell about such things from these data, since they do not give us the appropriate information.

Summary

In terms of statistical methodology, you have been introduced to the most important distribution in statistics – the normal distribution – and you have learned to use the distribution in two hypothesis tests, the one-sample and two-sample z-tests. In this unit, the normal distribution arose out of consideration of the sampling distributions of the sample mean: regardless of the distribution of the original data, such sampling distributions were seen to become more and more normal-like as the sample size, n, increased. You then learned about the normal distribution itself. You saw the way in which it depends on two quantities, the population mean, μ – controlling its location – and the population standard deviation, σ – controlling its spread. You also learned how any normal distribution can be related to a special normal distribution: the standard normal distribution with $\mu = 0$ and $\sigma = 1$. You then found that the sampling distribution of the sample mean can be approximated by a normal distribution with mean μ and standard deviation $\sigma/\sqrt{n}$, which is called the standard error of the mean.

The z-test was first introduced in its one-sample form to address null and alternative hypotheses concerning the value of μ. Its test statistic was developed in two forms, for σ assumed to be known and, more usefully, for σ unknown. You saw how the sampling distribution of the test statistic, and hence the critical values associated with the test, arose from the above results for the normal distribution. Having learned how to implement the one-sample z-test, you went on to learn how to adapt those ideas to produce the two-sample z-test; this is applicable to testing hypotheses concerning whether or not the means of two unrelated populations are equal. In each case, you applied what you learned about hypothesis testing in Unit 6 to interpret results in terms of the amount of evidence the data provide against the null hypothesis.

What you learned about childrens' reading abilities from the BCS survey has been summarised and discussed in Section 8.

Learning outcomes

After you have worked through this unit, you should be able to:

- appreciate the steps taken to make the unit's original question, which is rather general, more specific
- recall that the null and alternative hypotheses required for the z-test are expressed in terms of population means
- recognise a bell-shaped distribution
- appreciate that population distributions can have different shapes, some of which are normal
- appreciate that, whatever the shape of the population distribution, for a large enough sample size the sampling distribution of the mean is nearly always approximately normal
- appreciate the relationship between the location and spread of a normal distribution and its mean and standard deviation
- appreciate that it makes sense to think of normal distributions in terms of the number of standard deviations of the variable away from its mean, and that we can therefore think of all normal distributions in terms of only one distribution: the standard normal distribution
- apply the formula that transforms any variable x with a given normal distribution to the variable z with the standard normal distribution
- understand what is meant by the standard error (of the mean) and the estimated standard error in one- and two-sample situations
- write down the mean and standard deviation of the sampling distribution of the mean for samples of size n, given the population mean, μ, and standard deviation, σ
- follow the reasoning behind the one-sample z-test and apply the test when σ is assumed known
- adapt and apply the one-sample z-test when σ is unknown
- understand and apply the two-sample z-test to analyse the difference between means
- use Minitab to perform the one-sample z-test
- be aware of questions to ask which might lead to reservations about the conclusions of a hypothesis test
- be aware of some of the characteristics of the z-test, and recognise when it is necessary to exercise some caution in its use.

Solutions to activities

Solution to Activity 1

There are many possible answers to this question. You can test a child's reading ability by how well they read a coherent passage, recognise separate words, name letters, or pronounce separate words. Perhaps you have thought of other measures; or you may have thought in terms of a standard reading test of some kind.

Solution to Activity 2

Some of the factors you may have thought of are pre-school education, parents' education, precise age of child, whether there are other children in the family, mental or physical disability, social deprivation, quality of teaching, method of teaching, school class size, and parent's reading to the child at an early age. You may have been able to think of a different set of possibilities.

Solution to Activity 3

To be a random sample of exactly the sort you met in Unit 4, the sample would have had to be chosen by using random numbers to select children from a sampling frame of all 7- and 8-year-old children in the country. Clearly this was not done, so in this sense the sample is not random. However, you have previously met examples where a sample that was *not* chosen in this way was nevertheless considered to be representative in the same way that a formally selected random sample would be. In other words, the key question is not 'Was this sample chosen using a sampling frame and random numbers?', but 'Was this sample chosen in such a way that it has the same properties as one chosen using a sampling frame and random numbers?'

The answer to the second question is not so clear in this case. It might seem reasonable to treat the original BCS sample of people born in a particular week in 1970 as being representative of the general population of people born in Great Britain around that time, in the same way that a random sample would be representative. It is perhaps less reasonable to treat their 7- and 8-year-old children as if they were a random sample from the population of all 7- and 8-year-old children in 2004–2005. This is because in a true random sample of children, the ages of the children's parents would vary more – in this sample all the children have at least one parent born in a particular week in 1970. This might be quite a problem because the age and experience of their parents might well be linked to how a child's reading develops.

Solution to Activity 4

(a) The 7-year-old boys are identified in Table 1 by having a value of 1 in the third column (Gender – 1 denotes boy) and 1 in the fourth column (Coded age – 1 denotes 7 years old). There are six individuals in Table 1 that have 1 in each of the third and fourth columns. They have reading scores

$$106\ 110\ 134\ 25\ 172\ 160.$$

The sample size is $n = 6$.

(b) To calculate $\overline{x}$,

$$\sum x = 106 + 110 + 134 + 25 + 172 + 160 = 707,$$

and so

$$\overline{x} = \frac{\sum x}{n} = \frac{707}{6} \simeq 117.8.$$

Using Method 2 from Unit 3 (Subsection 3.1) to calculate s,

$$\begin{aligned}\sum(x-\overline{x})^2 &= \sum x^2 - \frac{(\sum x)^2}{n}\\ &= 97\,101 - \frac{(707)^2}{6}\\ &\simeq 13\,792.833.\end{aligned}$$

This means that the variance is

$$\begin{aligned}\frac{\sum(x-\overline{x})^2}{n-1} &= \frac{13\,792.833}{5}\\ &\simeq 2758.5667.\end{aligned}$$

So

$$\begin{aligned}s = \sqrt{\text{variance}} &= \sqrt{2758.5667}\\ &\simeq 52.5.\end{aligned}$$

Solution to Activity 5

The proportion of students on MS221 in the presentation in question achieving 75 marks is the actual number of students receiving 75 marks (21) divided by the total number of students sitting the exam (1234). That is,

$$\frac{21}{1234} \simeq 0.0170.$$

Solution to Activity 6

(a) (i) Sample mean $= \dfrac{15+35}{2} = \dfrac{50}{2} = 25.$

(ii) Sample mean $= \dfrac{65+77}{2} = \dfrac{142}{2} = 71.$

(iii) Sample mean $= \dfrac{65+52}{2} = \dfrac{117}{2} = 58.5.$

(iv) Sample mean $= \dfrac{37+80}{2} = \dfrac{117}{2} = 58.5.$

(b) The sample means of samples of size 2 are either integers (as in (i) and (ii) in part (a)) or else 'half-integers', that is, values of the form 'an integer plus a half' (as in (iii) and (iv) of part (a)).

Solution to Activity 7

The distribution of sample means of size 2 shown in Figure 3 is much smoother and less jagged than the distribution of the population data shown in Figure 2. The distribution of sample means of size 2 is also fairly symmetric, about a maximal value at around 70. However, there are slightly more sample means less than 70 than greater than 70, meaning that the distribution is slightly left-skew (see Subsection 5.2 of Unit 1). You might also note that the distribution fades away to almost nothing – corresponding to very rare sample mean values – at about 10 or so.

Solution to Activity 8

(a) Sample mean $= \dfrac{10+20+45}{3} = \dfrac{75}{3} = 25.$

(b) Sample mean $= \dfrac{82+24+33}{3} = \dfrac{139}{3} \simeq 46.3.$

(c) Sample mean $= \dfrac{52+61+73}{3} = \dfrac{186}{3} = 62.$

(d) Sample mean $= \dfrac{78+64+46}{3} = \dfrac{188}{3} \simeq 62.7.$

Solution to Activity 9

The distribution of sample means of size 3 shown in Figure 4 is much smoother than the distribution of sample means of size 2 shown in Figure 3 – it is made up of many more very short lines whose overall effect is closer to a smooth curve. The sampling distribution in Figure 4 is a little more compressed from side to side than that in Figure 3; that is, it has a smaller spread. The sampling distribution in Figure 4 is perhaps even closer to symmetric than the one in Figure 3. The maximum value about which the sampling distribution is approximately symmetric is, however, at approximately the same place as the maximum in Figure 3 – that is, at about, or a little under, 70. Finally, corresponding to its smaller spread, the distribution in Figure 4 fades away to almost nothing at about 20 or so (and just below 100).

Solution to Activity 10

The spread of the sampling distribution in Figure 5 is a little smaller again than the spread of the sampling distribution in Figure 4. It is also the case that any skewness apparent in Figure 4 is no longer apparent in Figure 5: this time, the distribution is symmetric, falling away smoothly on either side of a maximum value a little way below 70. But aside from the change in spread, the sampling distribution in Figure 5 is rather similar to the sampling distribution in Figure 4; in particular, the maximum is at approximately the same place in the two figures, while in both cases the sampling distributions fall away from the maximum, first more rapidly and then more slowly as they 'level out' a long way from the maximum.

Solution to Activity 11

As the sample size n increases, the sampling distributions, which all have the same symmetric shape, rise more and more sharply to a mode (at a little below 70, it seems). Also, the distributions become more and more compressed (i.e. the spread decreases as the sample size increases).

Solution to Activity 12

For $n = 2$, the sampling distribution of the mean is right-skew, but a little less so than the population distribution. As the sample size n increases, the sampling distributions again become more symmetric and bell-shaped. The distributions also become more and more peaked and compressed about the mode (at about £500).

Solution to Activity 13

The centre of this normal distribution is located at the value 1, so, as in Figure 11(b), this means that $\mu = 1$. The distribution also appears to have the same spread as the normal distribution in Figure 12(c), so $\sigma = 2$. To confirm these claims, notice that the x-axis labels on Figure 11(b) have 1 added to them (when $\mu = 1$) compared with the corresponding labels on Figure 11(a) (when $\mu = 0$); similarly, the x-axis labels on Figure 13 have 1 added to them (when $\mu = 1$) compared with the corresponding labels on Figure 12(c) (when $\mu = 0$).

Don't worry if you didn't get this activity right. There is much more on changing both μ and σ in the normal distribution in the Computer Book and Subsections 3.2 and 3.3 to follow.

Solution to Activity 14

(a) The mode of this normal distribution occurs at about $x = 10$. So $\mu \simeq 10$. Almost all the distribution is contained between $x = 4$ and $x = 16$ (i.e. within 10 ± 6). So $3\sigma \simeq 6$ and $\sigma \simeq 2$. That is, the normal distribution plotted in Figure 17 is approximately the normal distribution with mean $\mu = 10$ and standard deviation $\sigma = 2$.

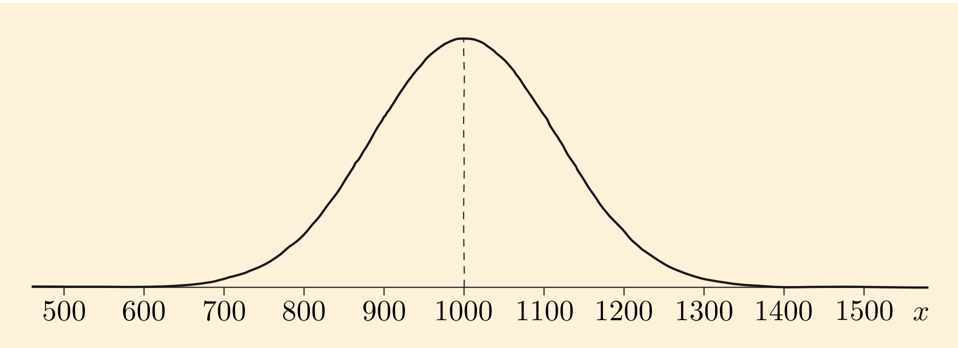
(b) The mode of this normal distribution occurs at about $x = 100$. So $\mu \simeq 100$. Almost all the distribution is contained between $x = 40$ and $x = 160$ (i.e. within 100 ± 60). So $3\sigma \simeq 60$ and $\sigma \simeq 20$. That is, the normal distribution plotted in Figure 18 is approximately the normal distribution with mean $\mu = 100$ and standard deviation $\sigma = 20$.

(c) The mode of this normal distribution occurs at about $x = 1$. So $\mu \simeq 1$. Almost all the distribution is contained between $x = 0.7$ and $x = 1.3$ (i.e. within 1 ± 0.3). So $3\sigma \simeq 0.3$ and $\sigma \simeq 0.1$. That is, the normal distribution plotted in Figure 19 is approximately the normal distribution with mean $\mu = 1$ and standard deviation $\sigma = 0.1$.

Solution to Activity 15

You should have obtained something like the sketches below, although, since you may have used different scales, yours could look a bit different. The important thing is that the information on your horizontal axes should match those in the figures.

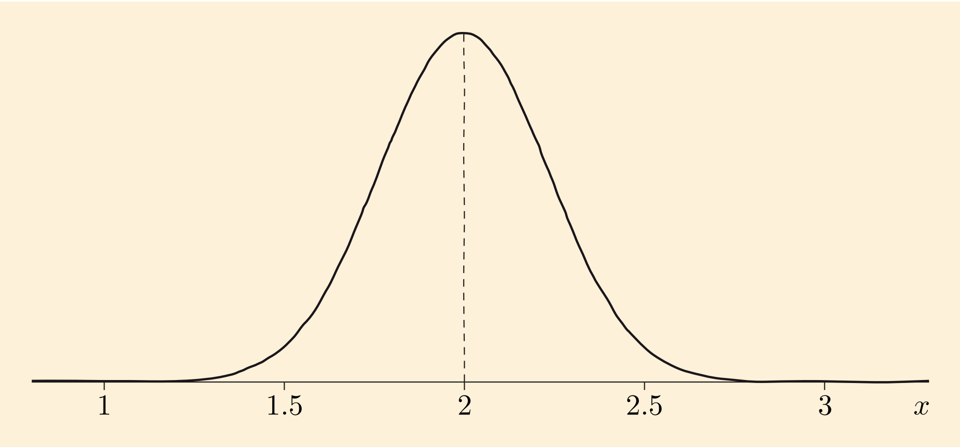
- The following normal distribution is centred at $\mu = 1000$ and has just about all the distribution contained within $1000 \pm (3 \times 100) = 1000 \pm 300$, i.e. between 700 and 1300.



The normal distribution with $\mu = 1000$, $\sigma = 100$

- The following normal distribution is centred at $\mu = 2$ and has almost all the

distribution contained within $2 \pm (3 \times 0.25) = 2 \pm 0.75$, i.e. between 1.25 and 2.75.



The normal distribution with $\mu = 2$, $\sigma = 0.25$

Solution to Activity 16

(a) Here $\mu = 10$ and $\sigma = 2$, so

$$z = \frac{x - 10}{2}.$$

(b) Here $\mu = 100$ and $\sigma = 20$, so

$$z = \frac{x - 100}{20}.$$

(c) Here $\mu = 1$ and $\sigma = 0.1$, so

$$z = \frac{x - 1}{0.1}.$$

If you prefer, you could equivalently write this as

$$z = \frac{x - 1}{1/10} = 10(x - 1).$$

Solution to Activity 17

(a) The appropriate formula is

$$z = \frac{h - \mu}{\sigma},$$

where $\mu = 1.75$ and $\sigma = 0.07$. Hence

$$z = \frac{h - 1.75}{0.07}.$$

(b) When $h = 1.96$,

$$z = \frac{1.96 - 1.75}{0.07} = \frac{0.21}{0.07} = 3.$$

So a height of 1.96 metres is 3 standard deviations above the mean height of 1.75 metres.

When $h = 1.61$,

$$z = \frac{1.61 - 1.75}{0.07} = \frac{-0.14}{0.07} = -2.$$

So a height of 1.61 metres is 2 standard deviations below the mean height of 1.75 metres.

When $h = 1.785$,

$$z = \frac{1.785 - 1.75}{0.07} = \frac{0.035}{0.07} = 0.5.$$

So a height of 1.785 metres is 0.5 standard deviations above the mean height of 1.75 metres.

You can check the picture of the distribution in Figure 14 (Subsection 3.2) to see if each of the values of h in this activity is the appropriate z standard deviations away from the mean.

Solution to Activity 18

(a) In each case the sampling distribution is symmetric about a mode at about 66 marks. So the means of the sampling distributions appear to be the same as the population mean $\mu = 66$ marks.

(b) The sampling distributions all look symmetric with a mode at about \$491 or so. So again the mean of each of the sampling distributions appears to be the same as the population mean $\mu = \$491$.

Solution to Activity 19

(a) The standard deviation of the sampling distribution of mean exam marks decreases (i.e. the distributions become more compressed) as the sample size n increases.

(b) The standard deviation of the sampling distribution of mean employees' earnings also decreases (i.e. the distributions become more compressed) as the sample size n increases.

Solution to Activity 20

(a) When $n = 25$,

$$\frac{\sigma}{\sqrt{n}} = \frac{22}{\sqrt{25}} = 4.4,$$

(b) When $n = 50$,

$$\frac{\sigma}{\sqrt{n}} = \frac{22}{\sqrt{50}} \simeq 3.11.$$

(c) When $n = 100$,

$$\frac{\sigma}{\sqrt{n}} = \frac{22}{\sqrt{100}} = 2.2.$$

Solution to Activity 21

When $n = 25$ and $\sigma = 0.01$,

$$\frac{\sigma}{\sqrt{n}} = \frac{0.01}{\sqrt{25}} = \frac{0.01}{5} = 0.002.$$

It follows that the sampling distribution of the mean for samples of 25 ball bearings from this manufacturer is approximately normal with mean $\mu = 2$ mm and standard deviation $\sigma/\sqrt{n} = 0.002$ mm.

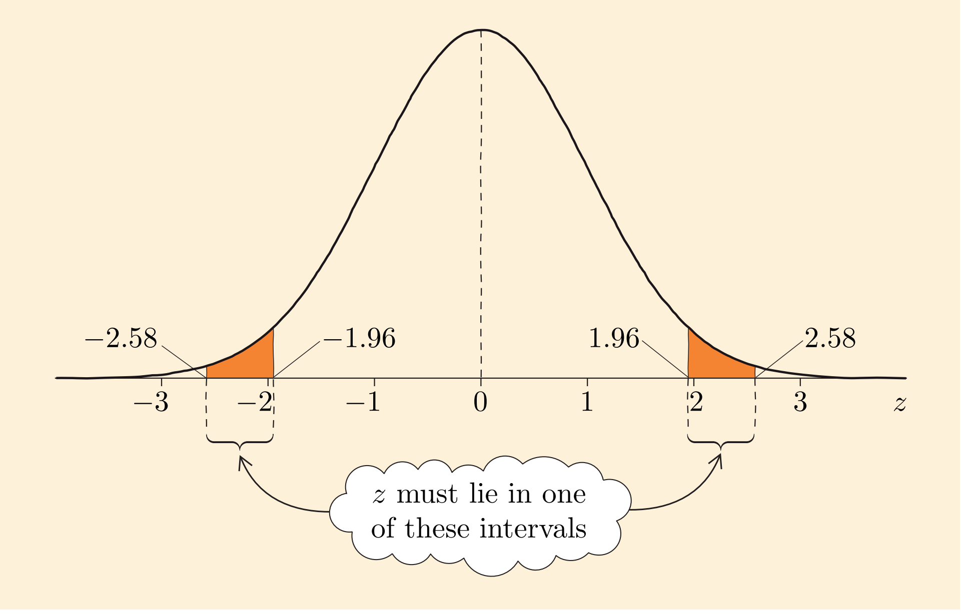
Solution to Activity 22

The value of z is

$$z = \frac{\overline{x} - A}{\text{SE}} = \frac{112 - 120}{15/\sqrt{100}} \simeq -5.33.$$

Solution to Activity 23

If H_0 is rejected at the 5% significance level but not at the 1% significance level, then z lies in the critical region shown in Figure 32 but not in the critical region shown in Figure 33, that is $1.96 \leq z < 2.58$ or $-2.58 < z \leq -1.96$. This is shown in the following figure.



Sketch of standard normal distribution with possible values of z indicated

Solution to Activity 24

(a) The null and alternative hypotheses are

$$H_0\colon \mu = 26.1$$
$$H_1\colon \mu \neq 26.1,$$

where μ is the population mean set-up time of the new method.

(b) $A = 26.1$, as this is the value of μ under H_0. The sample values are $\overline{x} = 20.9$ and $n = 53$.

(c) The test statistic is

$$z = \frac{\overline{x} - A}{\text{SE}} = \frac{20.9 - 26.1}{12.3/\sqrt{53}} \simeq -3.08.$$

(d) As -3.08 is less than -1.96 and -2.58, the null hypothesis is rejected at both the 5% significance level and the 1% significance level.

(e) There is strong evidence against H_0. Thus there is strong evidence that the mean set-up time under the new method differs from that under the old method – there is strong evidence that the new method is faster.

Solution to Activity 25

The appropriate null and alternative hypotheses are

H_0: $\mu = 116$

H_1: $\mu \neq 116$,

where μ is the population mean reading score of all British 8-year-old children in 2004–2005.

As the sample size, 283, is much greater than 25, it is appropriate to apply the z-test. We have

$$A = 116, \quad \overline{x} = 126.92, \quad n = 283, \quad s = 27.711.$$

The test statistic is

$$z = \frac{\overline{x} - A}{\text{ESE}} = \frac{\overline{x} - A}{s/\sqrt{n}} = \frac{126.92 - 116}{27.711/\sqrt{283}} \simeq 6.63.$$

The critical values are 1.96, -1.96 (5%) and 2.58, -2.58 (1%). Since $6.63 \geq 2.58$, we can reject the null hypothesis at the 1% significance level and conclude that there is strong evidence that the mean reading score of 8-year-olds in 2004–2005 was not equal to 116.

This might have been a little surprising if we had not seen a similar result for 7-year-old children in Example 7.

Solution to Activity 26

The null and alternative hypotheses are

H_0: $\mu = 381.50$

H_1: $\mu \neq 381.50$,

where μ is the population mean weekly wage (in \$) of female local government clerical officers and assistants in 2011.

As the sample size, 810, is greater than 25, it is appropriate to apply the z-test. We have

$$A = 381.5, \quad \overline{x} = 373.4, \quad n = 810, \quad s = 138.2.$$

The test statistic is

$$z = \frac{\overline{x} - A}{\text{ESE}} = \frac{\overline{x} - A}{s/\sqrt{n}} = \frac{373.4 - 381.5}{138.2/\sqrt{810}} \simeq -1.67.$$

The critical values are 1.96, -1.96 (5%) and 2.58, -2.58 (1%). Since $-1.96 < -1.67 < 1.96$, the null hypothesis is not rejected at the 5% significance level. There is little evidence that the mean weekly wage of female local government clerical officers and assistants differed from the overall mean weekly wage of female employees in 2011.

Solution to Activity 27

The null and alternative hypotheses are

H_0: $\mu_{\text{g}} = \mu_{\text{b}}$

H_1: $\mu_{\text{g}} \neq \mu_{\text{b}}$,

where μ_{g} and μ_{b} are the population mean reading scores for 8-year-old girls and boys, respectively. We have:

$\overline{x}_{\text{g}} = 127.49, \quad \overline{x}_{\text{b}} = 126.38, \quad n_{\text{g}} = 138, \quad n_{\text{b}} = 145,$

$s_{\text{g}} = 25.064, \quad s_{\text{b}} = 29.927.$

Both $n_{\mathsf{g}} = 138$ and $n_{\mathsf{b}} = 145$ are greater than 25, so we can assume that the z-test is applicable.

The estimated standard error is

$$\mathsf{ESE} = \sqrt{\frac{s_{\mathsf{g}}^2}{n_{\mathsf{g}}} + \frac{s_{\mathsf{b}}^2}{n_{\mathsf{b}}}} = \sqrt{\frac{25.064^2}{138} + \frac{29.927^2}{145}} \simeq 3.276,$$

and the test statistic is

$$z = \frac{\overline{x}_{\mathsf{g}} - \overline{x}_{\mathsf{b}}}{\mathsf{ESE}} \simeq \frac{127.49 - 126.38}{3.276} \simeq 0.34.$$

The critical values are $1.96, -1.96$ (5%) and $2.58, -2.58$ (1%). Since $-1.96 < 0.34 < 1.96$, we cannot reject the null hypothesis at the 5% significance level. There is no reason to doubt that the mean reading scores of 8-year-old girls and boys were the same.

Solution to Activity 28

Let 'E' denote quantities relating to children whose parental education ended by age 16, and 'C' denote quantities relating to children whose parental education continued after age 16. The null and alternative hypotheses are

$$H_0\colon \mu_{\mathsf{C}} = \mu_{\mathsf{E}}$$
$$H_1\colon \mu_{\mathsf{C}} \neq \mu_{\mathsf{E}},$$

where μ_{C} and μ_{E} are the population mean reading scores of interest. We have:

$$\overline{x}_{\mathsf{E}} = 116.12, \quad \overline{x}_{\mathsf{C}} = 123.15, \quad n_{\mathsf{E}} = 389, \quad n_{\mathsf{C}} = 199,$$
$$s_{\mathsf{E}} = 28.775, \quad s_{\mathsf{C}} = 24.603.$$

Both $n_{\mathsf{C}} = 199$ and $n_{\mathsf{E}} = 389$ are greater than 25, so we can assume that the z-test is applicable.

The estimated standard error is

$$\mathsf{ESE} = \sqrt{\frac{s_{\mathsf{C}}^2}{n_{\mathsf{C}}} + \frac{s_{\mathsf{E}}^2}{n_{\mathsf{E}}}} = \sqrt{\frac{24.603^2}{199} + \frac{28.775^2}{389}} \simeq 2.274,$$

and the test statistic is

$$z = \frac{\overline{x}_{\mathsf{C}} - \overline{x}_{\mathsf{E}}}{\mathsf{ESE}} \simeq \frac{123.15 - 116.12}{2.274} \simeq 3.09.$$

The critical values are $1.96, -1.96$ (5%) and $2.58, -2.58$ (1%). Since $3.09 \geq 2.58$, we can reject H_0 at the 1% significance level. There is strong evidence that the mean reading score of children whose parental education continued after age 16 differs from the mean reading score of children whose parental education ended by age 16. There is strong evidence that the children of parents who stayed longer in full-time education did better than those of parents who left education earlier.

Solutions to exercises

Solution to Exercise 1

(a) The 8-year-old children are identified in Table 1 by having a value of 2 in the fourth column (Coded age '2' denotes 8 years old). There are four individuals in Table 1 that have 2 in the fourth column. They have reading scores

$$118 \; 115 \; 56 \; 136.$$

The sample size is $n = 4$.

(b) To calculate $\overline{x}$,

$$\sum x = 118 + 115 + 56 + 136 = 425,$$

and so

$$\overline{x} = \frac{\sum x}{n} = \frac{425}{4} = 106.25.$$

To calculate s,

$$\begin{aligned}\sum (x - \overline{x})^2 &= \sum x^2 - \frac{(\sum x)^2}{n} \\ &= 48\,781 - \frac{425^2}{4} \\ &= 3624.75,\end{aligned}$$

which means the variance is

$$\frac{\sum (x - \overline{x})^2}{n - 1} = \frac{3624.75}{3} = 1208.25.$$

So

$$s = \sqrt{\text{variance}} = \sqrt{1208.25} \simeq 34.8.$$

Solution to Exercise 2

(a) The children of interest in this exercise are identified in Table 1 by having a value of 1 in the fifth column (Parental education '1' denotes finished aged 16 or less) and a value of 1 in the sixth column (Father's occupation '1' denotes managerial, technical, professional and skilled non-manual occupations). There are seven individuals in Table 1 that have 1 in both the fifth and sixth columns. They have reading scores

$$123 \; 110 \; 134 \; 110 \; 172 \; 136 \; 160.$$

The sample size is $n = 7$.

(b) To calculate $\overline{x}$,

$$\sum x = 123 + 110 + 134 + 110 + 172 + 136 + 160 = 945,$$

and so

$$\overline{x} = \frac{\sum x}{n} = \frac{945}{7} = 135.$$

To calculate s,

$$\begin{aligned}\sum (x - \overline{x})^2 &= \sum x^2 - \frac{(\sum x)^2}{n} \\ &= 130\,965 - \frac{945^2}{7} \\ &= 3390,\end{aligned}$$

which means the variance is

$$\frac{\sum(x-\overline{x})^2}{n-1} = \frac{3390}{6} = 565.$$

So

$$s = \sqrt{\text{variance}} = \sqrt{565} \simeq 23.8.$$

Solution to Exercise 3

Suitable null and alternative hypotheses are

H_0: For British children aged 8 in 2004–2005, the mean reading score for girls was equal to the mean reading score for boys

H_1: For British children aged 8 in 2004–2005, the mean reading score for girls was not equal to the mean reading score for boys.

Solution to Exercise 4

(a) For Population A, the six different samples of size 2 with their sample means are listed below:

$$\text{Sample: } 10\ 20; \text{ sample mean} = \frac{10+20}{2} = \frac{30}{2} = 15$$

$$\text{Sample: } 10\ 30; \text{ sample mean} = \frac{10+30}{2} = \frac{40}{2} = 20$$

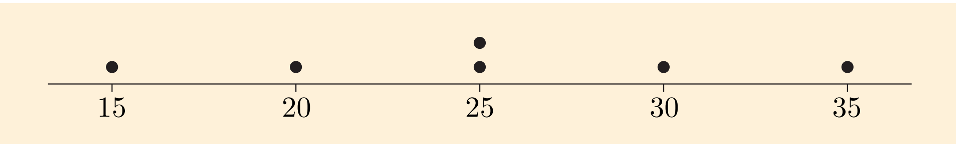
$$\text{Sample: } 10\ 40; \text{ sample mean} = \frac{10+40}{2} = \frac{50}{2} = 25$$

$$\text{Sample: } 20\ 30; \text{ sample mean} = \frac{20+30}{2} = \frac{50}{2} = 25$$

$$\text{Sample: } 20\ 40; \text{ sample mean} = \frac{20+40}{2} = \frac{60}{2} = 30$$

$$\text{Sample: } 30\ 40; \text{ sample mean} = \frac{30+40}{2} = \frac{70}{2} = 35$$

The sample means are plotted along the horizontal axis in the following figure.



Plot of values of sample means from Population A

(b) For Population B, the six different samples of size 2 with their sample means are listed below:

$$\text{Sample: } 10\ 38; \text{ sample mean} = \frac{10+38}{2} = \frac{48}{2} = 24$$

$$\text{Sample: } 10\ 39; \text{ sample mean} = \frac{10+39}{2} = \frac{49}{2} = 24.5$$

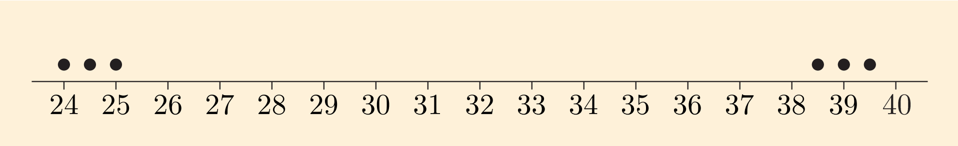
$$\text{Sample: } 10\ 40; \text{ sample mean} = \frac{10+40}{2} = \frac{50}{2} = 25$$

$$\text{Sample: } 38\ 39; \text{ sample mean} = \frac{38+39}{2} = \frac{77}{2} = 38.5$$

$$\text{Sample: } 38\ 40; \text{ sample mean} = \frac{38+40}{2} = \frac{78}{2} = 39$$

$$\text{Sample: } 39\ 40; \text{ sample mean} = \frac{39+40}{2} = \frac{79}{2} = 39.5$$

The sample means are plotted along the horizontal axis in the following figure.



Plot of values of sample means from Population B

(c) The points in the graph in part ((a)) are symmetrically distributed around a central mode, while the points in the graph in part ((b)) are split into two groups some distance apart. Hence the graph in part ((a)) seems more bell-shaped than the graph in part ((b)). This happens because the points in Population A are more symmetric – and more evenly spread out – than the points in Population B, which consist of three points close together (38, 39 and 40) and another far away (10).

Solution to Exercise 5

The distribution of reading scores – 'sample means' when $n = 1$ – is very jagged, but if you squint your eyes you get an impression of a fairly symmetric distribution with perhaps a slight preponderance of low, as opposed to high, values. The distribution of sample means of size $n = 2$ is smoother, though still with some jaggedness towards its right-hand side, fairly close to symmetric but with a little bit of left skewness. When $n = 3$ the distribution is smoother again, and any lack of symmetry is pretty small. It is also clear that the vertical scale of the sampling distribution of the mean when $n = 3$ is larger than the vertical scale of the distribution of the data $(n = 1)$. By the time $n = 10$, the distribution of sample means is very smooth, symmetric, bell-shaped/normal-like and with a larger vertical scale still.

So, again, we see that even though the population distribution is not especially normal-like, as the sample size n increases, the sampling distribution of the mean quite quickly becomes much more normal-like.

Solution to Exercise 6

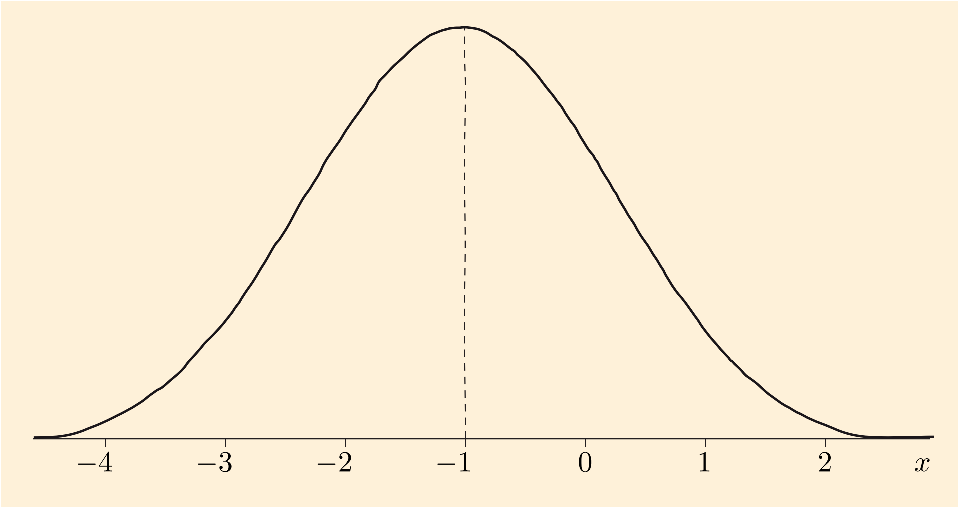
The mode of this normal distribution occurs at about $x = 10$. So $\mu \simeq 10$. Almost all the distribution is contained between $x = 0$ and $x = 20$ (i.e. within 10 ± 10). So $3\sigma \simeq 10$ and $\sigma \simeq 3.33$. That is, the normal distribution plotted in Figure 28 is approximately the normal distribution with mean $\mu = 10$ and standard deviation $\sigma = 3.33$.

Solution to Exercise 7

The mode of this normal distribution occurs at about $x = -5$. So $\mu \simeq -5$. Almost all the distribution is contained between $x = -12$ and $x = 2$ (i.e. within -5 ± 7). So $3\sigma \simeq 7$ and $\sigma \simeq 2.33$. That is, the normal distribution plotted in Figure 29 is approximately the normal distribution with mean $\mu = -5$ and standard deviation $\sigma = 2.33$.

Solution to Exercise 8

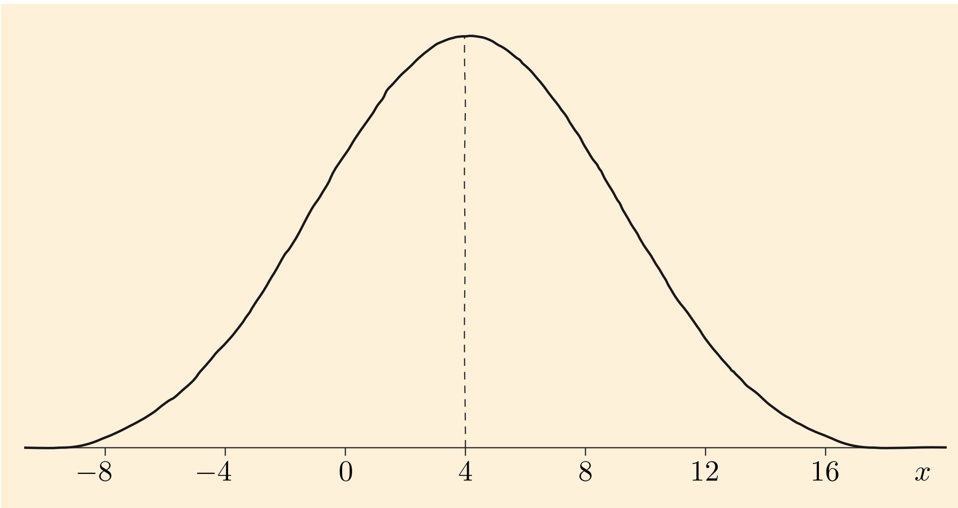
You should have obtained something like the sketch in the figure below, although, since you may have used different scales, yours could look a bit different. This normal distribution is centred at $\mu = -1$ and has just about all the distribution contained within $-1 \pm (3 \times 1) = -1 \pm 3$, i.e. between -4 and 2.



A normal distribution with $\mu = -1$, $\sigma = 1$

Solution to Exercise 9

You should have obtained something like the sketch in the figure below, although, since you may have used different scales, yours could look a bit different. This normal distribution is centred at $\mu = 4$ and has just about all the distribution contained within $4 \pm (3 \times 4) = 4 \pm 12$, i.e. between -8 and 16.



A normal distribution with $\mu = 4$, $\sigma = 4$

Solution to Exercise 10

(a) Here $\mu = 6$ and $\sigma = 3.3$, so

$$z = \frac{x-6}{3.3}.$$

(b) Here $\mu = -6$ and $\sigma = 2$, so

$$z = \frac{x-(-6)}{2} = \frac{x+6}{2}.$$

Solution to Exercise 11

The appropriate formula is

$$z = \frac{x-2}{10}.$$

When $x = 3$,

$$z = \frac{3-2}{10} = \frac{1}{10} = 0.1.$$

Solution to Exercise 12

The appropriate formula is

$$z = \frac{x-(-1)}{0.5} = \frac{x+1}{1/2} = 2(x+1).$$

When $x = 0$,

$$z = 2(0+1) = 2.$$

Solution to Exercise 13

(a) When $n = 9$,

$$\frac{\sigma}{\sqrt{n}} = \frac{283}{\sqrt{9}} = \frac{283}{3} \simeq 94.3.$$

(b) When $n = 25$,

$$\frac{\sigma}{\sqrt{n}} = \frac{283}{\sqrt{25}} = \frac{283}{5} = 56.6.$$

(c) When $n = 100$,

$$\frac{\sigma}{\sqrt{n}} = \frac{283}{\sqrt{100}} = \frac{283}{10} = 28.3.$$

Solution to Exercise 14

(a) When $n = 4$,

$$\frac{\sigma}{\sqrt{n}} = \frac{3.6}{\sqrt{4}} = \frac{3.6}{2} = 1.8.$$

(b) When $n = 19$,

$$\frac{\sigma}{\sqrt{n}} = \frac{3.6}{\sqrt{19}} \simeq 0.83.$$

(c) When $n = 300$,

$$\frac{\sigma}{\sqrt{n}} = \frac{3.6}{\sqrt{300}} \simeq 0.21.$$

Solution to Exercise 15

When $n = 40$ and $\sigma = 0.01$,

$$\frac{\sigma}{\sqrt{n}} = \frac{0.01}{\sqrt{40}} \simeq 0.0016.$$

It follows that the sampling distribution of the mean for samples of 40 one-litre bottles of water from this manufacturer is approximately normal with mean $\mu = 1.01$ litres and standard deviation $\sigma/\sqrt{n} = 0.0016$ litres.

Solution to Exercise 16

The appropriate null and alternative hypotheses are

H_0: $\mu = 96$

H_1: $\mu \neq 96$,

where μ is the population mean reading score of all British 7-year-old girls in 2004–2005.

As the sample size, $n = 190$, is much greater than 25, it is appropriate to apply the z-test. We have

$$A = 96, \quad \overline{x} = 113.42, \quad n = 190, \quad s = 25.464.$$

The test statistic is

$$z = \frac{\overline{x} - A}{\text{ESE}} = \frac{\overline{x} - A}{s/\sqrt{n}} = \frac{113.42 - 96}{25.464/\sqrt{190}} \simeq 9.43.$$

The critical values are $1.96, -1.96$ (5%) and $2.58, -2.58$ (1%). Since $9.43 \geq 2.58$, we can reject the null hypothesis at the 1% significance level. Hence there is strong evidence that the mean reading score of 7-year-old girls in 2004–2005 is not equal to 96.

This result corresponds to the similar result observed for all 7-year-old children (not just girls) in Example 7.

Solution to Exercise 17

The null and alternative hypotheses are

H_0: $\mu = 80$

H_1: $\mu \neq 80$,

where μ is the mean weight (in kg) of this breed of pig when fed the special diet.

As the sample size, 533, is greater than 25, it is appropriate to apply the z-test. We have

$$A = 80, \quad \overline{x} = 81.92, \quad n = 533, \quad s = 15.65.$$

The test statistic is

$$z = \frac{\overline{x} - A}{\text{ESE}} = \frac{\overline{x} - A}{s/\sqrt{n}} = \frac{81.92 - 80}{15.65/\sqrt{533}} \simeq 2.83.$$

The critical values are $1.96, -1.96$ (5%) and $2.58, -2.58$ (1%). Since $2.83 \geq 2.58$, the null hypothesis is rejected at the 1% significance level. There is strong evidence that the mean weight of this breed of pig when fed the special diet is not equal to 80 kg. There is strong evidence that the mean weight is higher for the special diet.

Solution to Exercise 18

(a) The null and alternative hypotheses are

H_0: $\mu = 4$

H_1: $\mu \neq 4$,

where μ is the population mean drying time in hours of the manufacturers' paint. As the sample size, $n = 40$, is greater than 25, it is appropriate to apply the z-test. We have

$$A = 4, \quad \overline{x} = 3.80, \quad n = 40, \quad s = 0.55.$$

The test statistic is

$$z = \frac{\overline{x} - A}{\text{ESE}} = \frac{\overline{x} - A}{s/\sqrt{n}} = \frac{3.80 - 4}{0.55/\sqrt{40}} \simeq -2.30.$$

The critical values are 1.96, -1.96 (5%) and 2.58, -2.58 (1%). Since $-2.58 < -2.30 \leq -1.96$, we can reject the null hypothesis at the 5%, though not at the 1%, significance level. We conclude that there is moderate evidence that the drying time given by the consumer magazine is incorrect. The manufacturers' paint appears to dry more quickly than the magazine claimed.

(b) You might think that such 'marginal' (moderate) evidence is not enough to conclude that the manufacturers' paint dries more quickly than the consumer magazine claimed.

The time at which paint is declared 'dry' is not well-defined: different customers might measure drying time differently or have different ideas about what 'dry' means.

Even if the measures are reliable and the test result is correct, 0.20 hours or 12 minutes is not a very large reduction. Most customers would not consider this an important difference.

You may have thought of other reservations.

Solution to Exercise 19

The null and alternative hypotheses are

H_0: $\mu_1 = \mu_2$

H_1: $\mu_1 \neq \mu_2$,

where μ_1 and μ_2 are the population mean reading scores of interest. Here and below, '1' denotes quantities relating to children with father's occupation coded 1 and '2' denotes quantities relating to children with father's occupation coded 2. The summary statistics are:

$$\overline{x}_1 = 120.55, \quad \overline{x}_2 = 117.17, \quad n_1 = 316, \quad n_2 = 203,$$
$$s_1 = 24.221, \quad s_2 = 30.085.$$

Both $n_1 = 316$ and $n_2 = 203$ are greater than 25, so we can assume that the z-test is applicable.

The estimated standard error is

$$\text{ESE} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{24.221^2}{316} + \frac{30.085^2}{203}} \simeq 2.513,$$

and the test statistic is

$$z = \frac{\overline{x}_1 - \overline{x}_2}{\text{ESE}} \simeq \frac{120.55 - 117.17}{2.513} \simeq 1.35.$$

(You might have got 1.34, correct to two decimal places, if calculating z all in one go. Such a difference doesn't matter.)

The critical values are $1.96, -1.96$ (5%) and $2.58, -2.58$ (1%). Since $-1.96 < 1.35 < 1.96$, we cannot reject H_0 at the 5% significance level. There is little evidence that the mean reading score of children whose father had an occupation coded 1 differed from the mean reading score of children whose father had an occupation coded 2. This goes against conventional wisdom, at least from other contexts.

Solution to Exercise 20

Let 'A' denote quantities relating to breast-fed babies and 'B' denote quantities relating to bottle-fed babies. The null and alternative hypotheses are

$$H_0\colon \mu_A = \mu_B$$
$$H_1\colon \mu_A \neq \mu_B,$$

where μ_A and μ_B are the population mean serum calcium levels of interest. The summary statistics are:

$$\overline{x}_A = 2.45, \quad \overline{x}_B = 2.30, \quad n_A = 64, \quad n_B = 169,$$
$$s_A = 0.292, \quad s_B = 0.274.$$

Both $n_A = 64$ and $n_B = 169$ are greater than 25, so we can assume that the z-test is applicable.

The estimated standard error is

$$\text{ESE} = \sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}} = \sqrt{\frac{0.292^2}{64} + \frac{0.274^2}{169}} \simeq 0.042,$$

and the test statistic is

$$z = \frac{\overline{x}_A - \overline{x}_B}{\text{ESE}} \simeq \frac{2.45 - 2.30}{0.042} \simeq 3.57.$$

(You might have got 3.56.)

The critical values are $1.96, -1.96$ (5%) and $2.58, -2.58$ (1%). Since $3.57 \geq 2.58$, we can reject the null hypothesis at the 1% significance level. There is strong evidence that the mean serum calcium level of week-old babies was different depending on whether they were breast-fed or bottle-fed. The evidence is that it was higher in those who were breast-fed.

Solution to Exercise 21

Let 'R' denote quantities relating to children from rural areas and 'U' denote quantities relating to children from urban areas. The null and alternative hypotheses are

$$H_0\colon \mu_{\mathsf{R}} = \mu_{\mathsf{U}}$$
$$H_1\colon \mu_{\mathsf{R}} \neq \mu_{\mathsf{U}},$$

where μ_{R} and μ_{U} are the population mean peak flow rates of interest (in litres per minute). The summary statistics are:

$$\overline{x}_{\mathsf{U}} = 226, \quad \overline{x}_{\mathsf{R}} = 231, \quad n_{\mathsf{U}} = 485, \quad n_{\mathsf{R}} = 637,$$
$$s_{\mathsf{U}} = 52, \quad s_{\mathsf{R}} = 53.$$

Both $n_{\mathsf{U}} = 485$ and $n_{\mathsf{R}} = 637$ are greater than 25, so we can assume that the z-test is applicable.

The estimated standard error is

$$\text{ESE} = \sqrt{\frac{s_{\text{R}}^2}{n_{\text{R}}} + \frac{s_{\text{U}}^2}{n_{\text{U}}}} = \sqrt{\frac{53^2}{637} + \frac{52^2}{485}} \simeq 3.160,$$

and the test statistic is

$$z = \frac{\overline{x}_{\text{R}} - \overline{x}_{\text{U}}}{\text{ESE}} \simeq \frac{231 - 226}{3.160} \simeq 1.58.$$

The critical values are 1.96, -1.96 (5%) and 2.58, -2.58 (1%). Since $-1.96 < 1.58 < 1.96$, we cannot reject the null hypothesis at the 5% significance level. Thus, there is little evidence to suggest that the mean peak flow rate differs between children who live in rural areas and those who live in urban areas.

Acknowledgements

Grateful acknowledgement is made to the following sources:

Subsection 1.1 photo (a reading class): Christopher Futcher / www.istockphoto.com

Subsection 1.2 figure (*Changing Britain, Changing Lives*): www.amazon.co.uk/Changing-Britain-Lives-Generations-Century/dp/0854736506

Subsection 1.2 figure (BAS 3 documentation): www.gl-assessment.co.uk
Section 2 photo (Big Ben), taken from: www.freewebs.com/mybellringing/famousbigbells.htm

Section 3 cartoon (statistics in Greece): www.causeweb.org

Subsection 3.2 figure ('abnormally normal or normally abnormal'), taken from: www.cduniverse.com/productioninfo.asp?pid=7362503

Subsection 5.2 ESE logo: www.citrus.k12.fl.us/ese/default.htm

Section 6 ESE logo ('a more relaxing form'): Easy Serving Espresso, www.easyservingespresso.co.uk

Section 8 photo (reserved sign) © Otnaydur / Dreamstime.com

Every effort has been made to contact copyright holders. If any have been inadvertently overlooked the publishers will be pleased to make the necessary arrangements at the first opportunity.

Index